Statistical image analysis and confocal microscopy

Alawadhi, Fahimah

Award date:
2001

Awarding institution:
University of Bath

Statistical Image Analysis and Confocal Microscopy

submitted by

Fahimah Alawadhi

for the degree of Ph.D.

of the

University of Bath

2001

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

Summary

Image analysis is an immensely active area, with a dramatic upsurge in fields such as medical image processing, zoological scans and computer vision.

In this thesis we consider data obtained from optical sections through an area of cartilage growth using fluorescence confocal microscopy. Our interests concern properties of cells in the specimen under study, such as their sizes, orientations, intensities and numbers. We adopt a *Bayesian approach* for *object recognition* incorporating *high level* prior models using a *marked point process* to handle the required geometrical features for the objects (cells). We use the recent developments in *Markov chain Monte Carlo* algorithms for varying dimension problems to explore the posterior distribution of cell configurations.

We work with the 2D optical section, then generalise the technique to handle the 3D problem using the consecutive optical sections provided. We use several techniques to simulate from the posterior distribution and to extract the required information for the cells under study. This is done, taking into consideration the efficiency, total computing cost, convergence and the mixing of the MCMC.

In order to extract the required aspects of the cells under study in a practical time we use a radical approach. We use a representative starting point instead of an extreme, blank configuration. The starting state should handle some initial information about the objects in the scene. To construct such state we developed an automated self-contained algorithm. We adopted some mathematical morphology operators, as well as some statistical tools to establish our approach.

Acknowledgements

First I would like to acknowledge and thank my supervisors, Professor Christopher Jennison and Dr. Merrilee Hurn for their invaluable intellectual support, help and encouragement during my research. They initiated many aspects in this thesis and provided much appreciated constructive advice regarding the process of completing my Ph.D.

I would like to thank Kuwait university for their scholarship which has allowed me to start and finish my Ph.D. at the University of Bath.

I am greatly indebted to Dr. Nick White and Dr. Rachel Errington from Department of Plant science at Oxford university for providing the data set for this thesis and for their valuable remarks and answers to my questions.

I am mostly grateful for Aaron Wilson who answered an endless number of computational questions and kept me going through the really dark times.

A general thank you goes to my friends, colleagues, computer support and staff members within the school of mathematical sciences at the university of Bath for being there when needed. Thanks for Kamala for taking care of my children whilst I was working at the university. I would like to mention Mona H. who have made my time in Bath so very enjoyable as well as Mona B. and May A.

I would specially dedicate many thanks to my parents for their unwavering support, my sister Shafeeqa for her guidance and encouragement that led me to start and finish my study. She has always had confidence that I would get to the end of the tunnel, even when I wasn't sure. Above all, I would like to dedicate my great fulness to my children Tareq, Mohammed and Farah.

Last but not least special thanks goes to my husband Farid whose love and patience (well most of times) impressed me while I and the children were away from him. Not only he supported me financially but also emotionally throughout this endeavour.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Outline

This thesis is organised as follows: In this chapter we shall discuss the general methodology for construction of Markov chains for Monte Carlo inference, particularly in a Bayesian context. Some aspects of prior modelling are considered. We focus on the two most common algorithms for implementing MCMC techniques. In chapter (2) we build up the general model for confocal microscope data. We then consider the prior model which we define a marked point process. We introduce a prior model that exhibits spatial interaction between objects in a scene.

Having identified the characteristics of this particular data set, we implement in chapter (3) the above methodology to construct the 2D target model. We then build up the MCMC algorithm and define the possible moves involved in the MCMC algorithm. We take into consideration the recent development of MCMC methods for multi-dimension state spaces which we discuss in detail. Consequently we test the 2D sampler performance. To speed up the convergence of the Markov chain we apply some modifications to the original algorithm as well as to the parameters of the target model. The latter modifications induce dynamic changes in the likelihood model. Finally we add the blurring effect to the model and test the performance of the sampler

accordingly.

As a result of the high dimension of the problem, there are several difficulties in actually reaching convergence when starting from extreme states such as the empty configuration. To avoid lengthy run-time we construct a starting value near the mode of the posterior model. Construction of such a starting state, either fully automated or by considering the use of expert knowledge, is carried out in chapter (4). We follow this by testing the performance of the model after using the appropriate techniques. The promising results of the 2D sampler after using a representative starting point lead us to image interpretation and extracting detailed information about the cells under study. This is handled in chapter (5) where we study some aspects of interest of the cells and apply statistical tools to help us in computing our estimates for those variables. We also check the robustness of these estimates to the prior assumptions.

In chapter (6) we consider the 3D data set using stacks of consecutive 2D optical sections. We generalise the 2D model to the 3D problem. The properties of the 3D records of the confocal microscopy are handled in detail as well as the identification criteria for the 3D objects. This is followed by describing the prior and the likelihood models and consequently the posterior model for the 3D problem. A brief implementation of the 3D MCMC algorithm is also included in chapter (6).

Chapter (7) discusses the 3D sampler performance using the original algorithm. In order to address the proposals of the chain toward the mode of the target model we apply some changes in the original algorithm. The performance of the improved sampler is again tested using extreme starting state as well as using a reliable starting point for the chain for the 3D problem.

In chapter (8) we apply radical approaches to explore the state space and overcome the mixing difficulties. We introduce two different techniques: the simulated tempering and a relaxed model approach. Finally, chapter (9) states the general conclusions and some questions raised throughout the thesis that motivate investigations and further studies.

1.2 Methodology

1.2.1 Introduction

In many applications of image analysis, the major concern is to describe the true scene X under view given an observed signal Y. We may wish to estimate X or a function $g(X)$. As an example, we may be interested in restoration of a satellite image of the earth's surface after removal of blur and noise, based on remotely sensed data Y. Suppose our concern is the proportion of the area classified as urban, agricultural, forest or water which in this example we label 1, 2, 3 and 4 respectively. The view under interest is partitioned into pixels. These pixels are labelled systematically $\{1,\ldots,n\}$, where n is the total number of pixels in the region. We define $X_i \in \{1,2,3,4\}$ to be the classification for pixel i, $i = 1,\ldots,n$. The proportions of interest are then $g_1(X), g_2(X), g_3(X)$ and $g_4(X)$ where

$$g_j(X) = \sum_{i=1}^{n} \frac{I[X_i = j]}{n}, \quad j = 1,2,3,4.$$

The remotely sensed data Y are actually degraded signals for the underlying scene X which are recorded for a subset S of the pixels, $Y = (Y_1,\ldots,Y_S)$. In many cases the received signals are a convolution or blurred version of the true scene with sensor noise resulting from the recording process. In the presence of the blur effect the blurred value Y at pixel i has expectation equal to a weighted average of X_i and the X values of the pixels around it. Given blurred records, if the blurred values across the scene are assumed to follow the same pattern of weight values then we need all pixels that may contribute to these records. That is we need to enlarge n such that the pixels lying at the edge of the region have the necessary adjacent pixels, so we choose $S < n$. When there is no blurring effect then the dimensions of the data vector Y and the output vector X are the same. The above example is considered as a *low level image* or pixel-wise problem where for each pixel i in the scene we wish to estimate a corresponding

output X_i.

Image analysis also incorporates *high level image* problems where the output deals with more complex global features of the image. In high level image models, with which we shall be working in this study, we specify the image as a collection of objects, each spanning a substantial number of pixels. Object recognition is a task that is considered a high level image problem. For example in this study the records Y are obtained from confocal microscopy for an area of cartilage growth. The image consists of an unknown number of cells. We are interested in both the number of cells, and in some properties of these cells. For illustration suppose the shape of the cells can be approximated by discs with fixed known radii and we are interested in the location of the centres of the cells falling within the image. Let N denote the number of cells in the specimen and $X = (X_1, \ldots, X_N)$ represent the N cell centres, then each X_i is a two dimensional pair representing the Cartesian coordinate of the disc centre.

Another important example of a high level problem is an image containing a mixture of cell types. For instance we may expand the exposition of the previous image example such that the objects may be approximated by circles of different fixed radii. We may be interested in estimating the number of cells of each type, $N_i, i = 1, 2, \ldots, m$, where m is the total number of classes of the cells, or perhaps we are interested in the location of each cell in the scene and its type. The output configuration X is then an unordered set $\{X_1, \ldots, X_N\}$ and each $X_i = ((c_x, c_y), tp)$, where (c_x, c_y) is the Cartesian coordinate of the $\underline{ith}$ cell centre and $tp \in \{1, \ldots, m\}$ represents the type of the cell i.

1.2.2 Bayesian inference

The records $Y = (Y_1, \ldots, Y_S)$ can be considered as convolutions of the sensor noise with the blurred values of the underlying scene. The form of the likelihood function $\mathcal{L}(y|x)$ of the observed record $Y = y$ depends on the method of imaging. We shall assume that given a particular true scene $X = x$, the random variables $Y_1, \ldots, Y_S$ are conditionally independent. Moreover, assume that each Y_i has a known conditional density function

$\pi_{Y_i|X}(y_i|x)$. Thus the conditional density of the observed records Y given $X = x$ is simply

$$\mathcal{L}(y|x) = \prod_{i=1}^{S} \pi_{Y_i|X}(y_i|x).$$

Using direct observed data, each record Y_i depends only on the pixel i it is representing: its intensity value, say λ_i is a function only of x_i in the low-level pixel description. Therefore the conditional density $\pi_{Y_i|X}(y_i|x)$ in this case is just $\pi_{Y_i|X}(y_i|x_i)$. In the high level model the interpretation of the output representation differs from that of the low level model. Each X_j there represents an object covering several pixels in the scene. If we let $x_{j(i)}$ be the object covering pixel i, then $x_{j(i)}$ implies the intensity of pixel i. Therefore the conditional density $\pi_{Y_i|X}(y_i|x)$ for the high level model using the direct observation can be written as $\pi_{Y_i|\lambda_i}(y_i|\lambda_i)$ or more generally as $\pi_{Y_i|X_{j(i)}}(y_i|x_{j(i)})$. On the other hand under the blurring effect this conditional density depends not only on a single unit x_i or $x_{j(i)}$ but on a set of units around pixel i. Let this set be denoted by Λ_i, then the conditional densities of the pixel image model can be expressed as $\pi_{Y_i|\Lambda_i}(y_i|\Lambda_i)$. In the high level model the conditional density of the record Y_i will depend on X_j, the object covering pixel i and the intensities of the adjacent pixels, which may belong to other objects.

Let $\mathbb{E}$ be the sample space of the configuration $X = (X_1, \ldots, X_n)$ where $\mathbb{E}$ is $\mathbb{R}^n$ or any suitable state space depending on the application. For example in the one type object problem defined on section (1.2.1), the sample space for each object X_i which consists merely of a location is $\mathbf{U}$. Here the sample space $\mathbf{U}$ of an object is the window or the scene under study, whereas the sample space of the whole configuration X is $\mathbb{E} = \cup_{n=0}^{\infty} \mathbf{U}^n$. On $\mathbb{E}$ we define a prior distribution for X. This prior distribution should be chosen to capture general knowledge about the image.

Low level models are commonly provided by Markov Random Field (MRF) models for the set of pixel values. In these models, the conditional distribution of one pixel value X_i given the rest of the image X depends only on the values X_j at a small number

of neighbouring pixels. Such models support the belief that pixels which are close in the grid take values which are similar. The Markov models can also be used in high level modelling, but they are no longer pixel-based. Rather, they are object based. The choice of a suitable model in high level tasks is related to the task being performed. In the object recognition problems a widely used prior density class is that of *pairwise interaction models* which exhibit repulsion or inhibition between neighbouring objects.

To combine the two main parts of the problem, the likelihood of Y and the prior distribution of X we follow the Bayesian approach. We shall use the posterior distribution $\pi_{X|Y}$ of the scene X given the data Y, to make inference about the underlying scene X given Y where

$$\pi_{X|Y}(x|y) \propto \mathcal{L}(y|x)\pi_X(x).$$

Under a known posterior distribution $\pi_{X|Y}$ the expectation of $g(X)$, the function under interest, will be given by

$$E_{\pi_{X|Y}}(g(x)|Y = y) = \int_{x\in\mathbb{E}} g(x)\pi_{X|Y}(dx|y). \tag{1.1}$$

1.2.3 Simulation techniques

In image analysis problems, X is high dimensional and evaluation of $\pi_{X|Y}$ is a challenging problem. Analytical calculation of (1.1) is not possible because of the complexity of the posterior distribution and the fact that this is only known up to proportionality. Simulation inference can be used instead. The technique of simulation is widely used in statistical inference e.g. Ripley (1987). Suppose, for example, that we have a way to obtain independent samples $x^{(1)}, x^{(2)}, \ldots$ from $\pi_{X|Y}$, we could then approximate the expectation of (1.1) by the *empirical estimate*

$$\overline{g}_t = \frac{1}{t}\sum_{i=1}^{t} g(x^{(i)}), \tag{1.2}$$

By the strong law of large numbers as $t \to \infty$,

$$\overline{g}_t \xrightarrow{a.s.} E_{\pi_{X|Y}}(g).$$

Typically the distribution $\pi_{X|Y}$ is too complex for direct simulations largely because the normalising constant is unknown. Thereupon the indirect approach of *Markov Chain Monte Carlo*, MCMC, must be applied. This approach will simulate correlated samples $\{X^{(i)}\}$ from $\pi_{X|Y}$. We need not worry about the dependence of the samples if they are drawn in the correct proportion and for a sufficient large sample size.

1.2.4 Markov chain Monte Carlo methods

Markov chain Monte Carlo methods are widely advocated in a variety of situations where the complexity of the distribution of interest is an issue. In these situations usually the direct sampling from such a complicated models is not possible. The key idea of the Markov chain Monte Carlo methods is to generate an iterative sequence of samples in such away that it converges in distribution to the model of interest. To implement this strategy many attempts were made to define algorithms for constructing chains with specified equilibrium distributions. The most common, well-known algorithms for constructing chains with specified equilibrium distributions were defined by Metropolis *et al.* (1953), Hastings (1970) and Geman and Geman (1984). A wide range of discussion papers on MCMC theory and application can be referred to, for example, Besag and Green (1993), Smith and Roberts (1993), Tierney (1994), Besag *et al.* (1995) and Gilks *et al.* (1996). In this section, we shall briefly discuss in an appropriate framework the theory of the MCMC technique.

To sample from a specified distribution π on $\mathbb{E}$, we construct a Markov chain transition kernel $\mathbb{P}(x, A)$. The transition kernel $\mathbb{P}$ is a map, $\mathbb{P} : \mathbb{E} \times \mathbb{E} \to [0, 1]$, that implies the target distribution π is a stationary distribution of the chain. The distribution of

$X^{(t+1)}$ given $X^{(t)}$ satisfies

$$P(X^{(t+1)} \in A | X^{(0)} = x^{(0)}, \ldots, X^{(t-1)} = x^{(t-1)}, X^{(t)} = x) = \mathbb{P}(x, A).$$

We say that π is the invariant measure (hence equilibrium) of the Markov chain if it satisfies the *general balance* equation

$$\int_{x \in \mathbb{E}} \pi(dx)\mathbb{P}(x, A) = \pi(A), \qquad \text{for all measurable sets A} \subset \mathbb{E}. \tag{1.3}$$

General balance, $\pi\mathbb{P} = \pi$, is also referred to as the *global balance.*

The conditional distribution of $X^{(t)}$ given $X^{(0)} = x^{(0)}$ is

$$P(X^{(t)} \in A | X^{(0)} = x^{(0)}) = \mathbb{P}^t(x^{(0)}, A),$$

where $\mathbb{P}^t$ denotes the kernel $\mathbb{P}$ after iterating it t times. The following theorem proved by Nummelin (1984) states conditions on the transition kernel $\mathbb{P}$ under which the distribution of the chain converges to the probability distribution π as $t \to \infty$.

Theorem 1 *Suppose $\mathbb{P}$ is $\pi-$irreducible and $\pi\mathbb{P} = \pi$. Then $\mathbb{P}$ is positive recurrent and π is the unique invariant distribution of $\mathbb{P}$. If $\mathbb{P}$ is also aperiodic, then, for $\pi-$almost all x;*

$$\|\mathbb{P}^t(x, \cdot) - \pi\| \to 0, \tag{1.4}$$

where $\|\cdot\|$ denotes the total variation distance. In other words if $\mathbb{P}$ satisfies the conditions stated in theorem (1), then for all measurable sets $A \subset \mathbb{E}$ and for almost all $x \in \mathbb{E}$

$$\lim_{t \to \infty} \mathbb{P}^t(x, A) = \pi(A).$$

A chain is $\pi-$*irreducible* if starting at any initial state $x \in \mathbb{E}$, then for all measurable sets A $\subset \mathbb{E}$ with $\pi(A) > 0$ there exists $t > 0$ such that $P(X^{(t)} \in A | x^0 = x) > 0$. The chain is *aperiodic* if the chain does not oscillate between different sets of spaces in a

regular periodic movement. The term *positive recurrent* is defined as follows: let τ_A be the first return time to state $A \subset \mathbb{E}$ where $\pi(A) > 0$, then we say the $\pi-$irreducible chain $X^{(t)}$ is recurrent if $P(\tau_A < \infty) = 1$ and is positive recurrent if $E(\tau_A) < \infty$.

If the chain is $\pi-$irreducible, aperiodic, positive recurrent and if the initial value of $X^{(0)}$ is sampled from π, then all subsequent iterations using MCMC will also be distributed according to π.

Theorem (1.4) ensures, for large t, the stationary distribution π is the limiting distribution of the Markov Chain regardless of the starting value of the chain. This point is very helpful in practice because, in general, sampling the initial state $X^{(0)}$ directly from $\pi(\cdot)$ is not feasible. Suppose this is true and we run the chain from a fairly arbitrary state in $\mathbb{E}$ for a long time. Taking a sufficiently large number of correlated samples should produce the same results as a given number of independent ones. To estimate the properties of the target distribution we want the chain to be invariant of its starting state so we neglect the samples obtained up to the *burn-in time.* Allowing a sufficient burn-in time before collecting samples ensures that $X^{(0)}$, the first X we use, has a distribution rather close to π. Then the simulated values of the chain can be used to summarise the features of $\pi(\cdot)$.

An easier condition to check than global balance is the *detailed balance* equation or time reversible condition which is given by

$$\pi(dx)\mathbb{P}(x, dx') = \pi(dx')\mathbb{P}(x', dx), \quad \forall x, x' \in \mathbb{E}. \tag{1.5}$$

The detailed balance condition, if confirmed, will ensure general balance condition as follows

$$\begin{aligned} \int_{x\in\mathbb{E}} \pi(dx)\mathbb{P}(x, A) &= \int_{x\in\mathbb{E}} \pi(dx) \int_A \mathbb{P}(x, dx') \\ &= \int_{x\in\mathbb{E}} \int_A \pi(dx)\mathbb{P}(x, dx') \end{aligned}$$

$$= \int_{x\in\mathbb{E}} \int_A \pi(dx')\mathbb{P}(x', dx)$$

$$= \int_A \pi(dx') \int_{x\in\mathbb{E}} \mathbb{P}(x', dx)$$

$$= \pi(A).$$

The above results are verified for a time homogeneous transition kernel. It may, however, be convenient to define several transition types and repeat these according to a specified pattern. One example is where a separate transition type is used to update each pixel element of an image X. Another possibility, in high level models, is to define separate transitions for adding, deleting or modifying objects in the scene. If, at each step, a transition type is chosen at random from the class of transitions, we have a time homogeneous process once again. Suppose instead we choose a pattern of transition types and then repeat it accordingly. For example we may update the components of $X^{(t)} = (X_1^{(t)}, \ldots, X_n^{(t)})$ individually from 1 through to n. Or we may set up the sampler scheme to be a systematic cycle through the available moves. The key point is to think about the full cycle of transitions as a single sampler step. Let $\mathbb{P}_i$ be the transition kernel for implementing transition type i, $i \in \mathcal{C}$, where $\mathcal{C}$ is the defined transition types. Suppose we cycle through $i_1, i_2, \ldots, i_k$ repeatedly. Then the transition kernel $\mathbb{P}$ for the sampler step is the product of the individual kernels, $\mathbb{P}_{i_1}\mathbb{P}_{i_2}\ldots\mathbb{P}_{i_k}$.

For each individual transition type i it is assumed that detailed balance is maintained, which implies that $\pi\mathbb{P}_i = \pi$, $\forall i$. Therefore general balance is satisfied for the sampler step,

$$\pi\mathbb{P} = \pi\mathbb{P}_1\ldots\mathbb{P}_n = \pi$$

As we defined a single step in the new chain to be the cycle of n sequential steps, then by the standard theory $\{X^{(n)}, X^{(2n)}, \ldots\}$ has limiting distribution π. But because general

balance is satisfied for the individual steps then in the long run π will, approximately, be the limiting distribution of the intermediate steps as well. Suppose

$$\|\mathbb{P}(x^{(0)}, \cdot) - \pi(\cdot)\| < \epsilon$$

then

$$\begin{aligned} \|\mathbb{P}\mathbb{P}_1(x^{(0)}, \cdot) - \pi(\cdot)\| &= \|\mathbb{P}\mathbb{P}_1(x^{(0)}, \cdot) - \mathbb{P}_1\pi\| \\ &= \|(\mathbb{P} - \pi)\mathbb{P}_1\| \\ &< \epsilon, \end{aligned}$$

and as $\epsilon \to 0$ then $\mathbb{P}\mathbb{P}_1 \to \pi$. It should be noticed that even if there is detailed balance for each $\mathbb{P}_i$ that does not imply detailed balance for $\mathbb{P}$. The detailed balance of the sampler step is attained if the systematic cycle is reversible. A reversible cycle is produced by defining our cycle through $i_1, i_2, \ldots, i_k, i_k, i_{k-1}, \ldots, i_1$ repeatedly.

Detailed balance is an easy condition to check, and one which implies general balance. We shall demonstrate this idea for specific types of MCMC algorithms in the next section.

The Metropolis-Hastings algorithm

This algorithm was first proposed by Metropolis *et al.* (1953) and extended by Hastings (1970). The algorithm is designed to give samples from a distribution π. It defines a proposal kernel $q(x, \cdot)$ to produce a potential new state $x' \in \mathbb{E}$. The proposed candidate x' is accepted with probability α where

$$\alpha(x, x') = \min\left\{1, \frac{\pi(x')q(x', x)}{\pi(x)q(x, x')}\right\}. \tag{1.6}$$

If we are currently at time t and x' is accepted, then $X^{(t)} = x'$ otherwise the chain does not move, i.e., $X^{(t)} = x^{(t-1)}$. Metropolis *et al.* (1953) had proposed a symmetric proposal density q which simplifies the acceptance probability α to

$$\alpha(x, x') = \min\left\{1, \frac{\pi(x')}{\pi(x)}\right\}. \tag{1.7}$$

Formally the target distribution π is defined with respect to a $\sigma-$finite measure. The proposal density q could be defined with respect to a different $\sigma-$finite measure from that for π. For example, if only one element of $x = (x_1, \ldots, x_n)$, $n > 1$ is updated at a time. We may deal with a more complicated case when parts of the sample space $\mathbb{E}$ have X of different dimensions. If x and x' are defined in different dimension spaces, then in this case even $q(x, x')$ and $q(x', x)$ will be defined with respect to different measures. A more detailed analysis of these cases will be verified in section (3.4).

The transition kernel $\mathbb{P}(x, x')$ using the Metropolis-Hastings algorithm can be written as

$$q(x, x')\alpha(x, x') \text{ if } x' \neq x.$$

The probability of the algorithm remaining at x is

$$r(x) = 1 - \int_{y \neq x} q(x, dy)\alpha(x, y).$$

Using the definition of α in (1.6) and of $\mathbb{P}$ it can be easily shown that the algorithm attains detailed balance and hence the general balance.

The choice of the distribution $q(\cdot, \cdot)$ is arbitrary provided that $q(x, x') > 0$ if and only if $q(x', x) > 0$. It is convenient to choose a q that is simple and fast to sample from and for which it is easy to evaluate the acceptance probability. However the relation between q and $\pi(\cdot)$ will affect the rate of convergence.

In implementing the algorithm with single-site updating just one element of X is changed at each step. The sampler may update one randomly picked element among

the scene, X_i, $i = \{1, \ldots, n\}$ and consider this as one iteration whether the proposal element is accepted or not. Thus the transition kernel $\mathbb{P}$ is time homogeneous. Alternatively, we can update all components of X in a raster order. In this case we consider the sequential updating of all the components as a single iteration. The transition kernel $\mathbb{P}$ is the product of the individual transition kernels, i.e., $\mathbb{P} = \mathbb{P}_1\mathbb{P}_2 \ldots \mathbb{P}_n$, where $\mathbb{P}_i$ is the transition kernel for updating component i.

The Gibbs Sampler

The Gibbs sampler was given its name by Geman and Geman (1984) who used it for analysing Gibbs distributions on a lattice. The algorithm constructs the transition kernel $\mathbb{P}$ using the full conditional densities of each component X_i, $i = 1, \ldots, n$, given the values of the other components $X_{-i} = \{X_j; j \neq i, j = 1, \ldots n\}$. We denote this density by $\pi_{X_i|X_{-i}}(x_i|x_{-i})$. Suppose we are at time t and want to update the chain, then (as it is with the Metropolis-Hastings algorithm) we either use a random sampler or a systematic scan sampler. At each iteration, the random sampler picks a random component say, X_i, $i \in \{1 \ldots n\}$ to update, then the conditional density for $X_i^{(t)}$ becomes $\pi_{X_i|X_{-i}}(x_i|X_{-i} = x_{-i}^{(t-1)})$. In the systematic scan, we update all the components in turn during one iteration using the marginal conditional densities of the components. In progressing from $X^{(t-1)}$ to $X^{(t)}$, the value of X_i is obtained by sampling from

$$\pi_{X_i|X_{-i}}(x_i|x_1^{(t)}, \ldots, x_{i-1}^{(t)}, x_{i+1}^{(t-1)}, \ldots, x_n^{(t-1)}).$$

Hence, to update X we make random draw from these full conditional densities for each of its components. The iteration is completed when all the components are updated. Hence, the transition probability from $x^{(t-1)}$ to $x^{(t)}$ is given by

$$\mathbb{P}(x^{(t-1)}, x^{(t)}) = \prod_{l=1}^{n} \pi_{X_l|X_{-l}}(x_l^{(t)}|x_1^{(t)}, x_2^{(t)}, \ldots, x_{l-1}^{(t)}, x_{l+1}^{(t-1)}, \ldots, x_n^{(t-1)}).$$

The Gibbs sampler can be regarded as a special case of Metropolis-Hastings algo-

rithm in which the acceptance rate α is one, meaning that the candidate x' is always accepted. Consequently the proposal kernel $q(\cdot, \cdot)$ is also the transition kernel $\mathbb{P}$. The results hold for the chain to attain detailed balance.

Chapter 2

Building up the target model

2.1 The noise model

2.1.1 Description of the confocal microscope data

The experimental data consist of an image Y collected by confocal fluorescence microscopy for an area of cartilage growth. In order to tackle the degradation resulted from imaging using confocal microscopes, we shall briefly describe the methodology of confocal microscopy.

A confocal microscope is a device used to image tissue in three dimensions. In essence a confocal microscope works by first staining the specimen under study with a fluorescent dye. Then by using a system of lenses and pinholes a laser beam is focused at a point in the pattern and a record is made of the fluorescence coming from that point. It images one object point at a time to provide a good segmentation of that point. The record is made by counting the number of photon hits to the receiving lens (see figure $(2-1)$). The scanning is achieved by repeating this process across a regular lattice of points in the specimen. An image of the entire specimen at certain depth is obtained by moving the focal point in different directions. The record obtained for a single point is then scaled to be in the range of $\{0, 1, \ldots, 255\}$ using mostly linear transformations. The three dimensional representation of the specimen is acquired

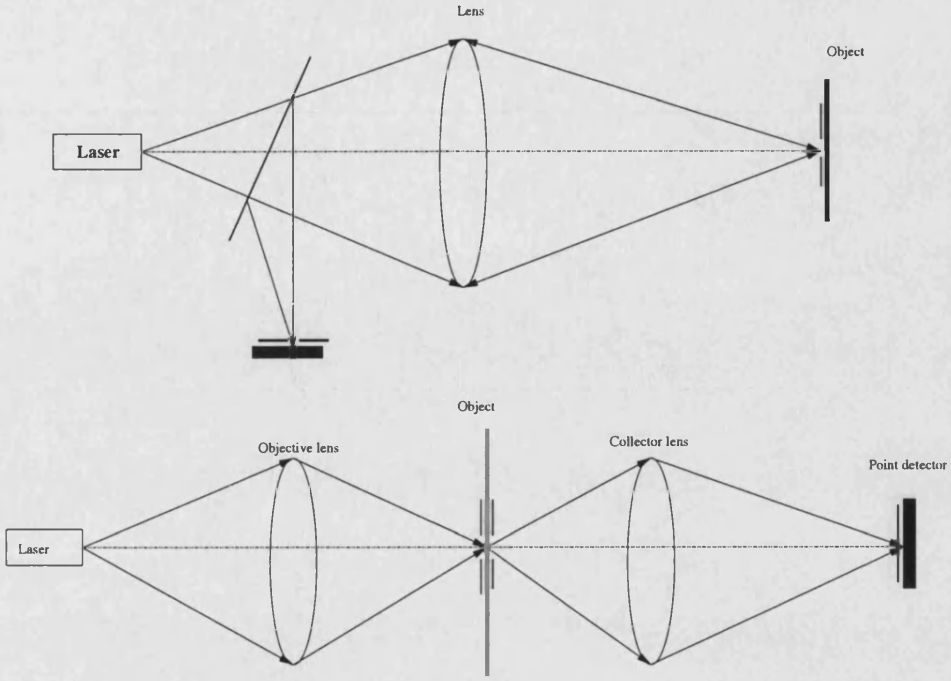

Figure 2-1: Typical arrangements of a scanning optical microscope.

by scanning consecutive sections with a change of focus setting between each section (optical serial sectioning).

Fluorescence microscopy with dye-treated samples gives a stronger signal than the conventional microscopes which do not use dyes. The intensity of the light transmitted from any point in the specimen depends on the reflective index of the material. Some materials emit low signals therefore fluorescent dye is used to improve the weak signal levels. The primary radiation, which is usually the laser beam, excites fluorescence which is then imaged using suitable filters. These filters aim to detect only fluorescent radiation so to produce an incoherent fluorescent field proportional to the intensity of incident radiation.

A digital image of the specimen can consist of up to 1024×1024 pixels with 256 grey levels for each optical slice or section through the specimen in the focal plane of the microscope. The optical system along with the scaling transformation of the

records gives space invariant imaging which ensures that the resolution and contrast are identical across one entire field of view. To modify the resolution of the image formation and to increase contrast in the image some techniques are used such as *dark ground* technique. Usually in the specimen being imaged the background media emit no, or very weak, signals. In the dark ground technique these values are increased to provide better contrast. The increment is also applied to all other parts of the specimen so as to form an image that is space invariant.

In confocal microscopy a common problem with recording is the low *signal-to-noise* ratio due to photon quantum noise. The microscope counts a small number of photons detected over time. Assuming these counts have a Poisson distribution then there is Poisson variation for these counts. The low signal-to-noise ratio is also due to other major reasons. One is that although in theory the microscopy attempts to capture only the fluorescence emitted from the focal point of the incident laser, in practice it may not eliminate out of focus information. In a confocal microscope the light from the point source probes not only a single point itself but a larger region. In a 2D image it is approximated as a disc region with different diameters depending on the microscope, and in 3D as a cylindrical shape. Hence, a nominal point of exposure light will give rise to an image density distribution which is called the *point spread function.* The scattering feature is for both the incoming laser and the outgoing fluorescent light. This can be treated as a blurring effect and we shall do so in later sections of this chapter. The other reasons for the low signal to noise ratio are the scaling effects for the image and the effects of diffraction and attenuation of the incoming laser and outgoing fluorescence lights of the object being studied. As a result of the above, optical sections that are close to the surface usually have relatively clear signals and less blurring, whereas the signals received from the deeper sections are more attenuated and more blurred.

2.1.2 The likelihood function

The data consist of an image $Y = (Y_j; j = 1, \ldots, S)$ with 256 grey levels. To analyse the data statistically we define c_j, $j = 1, \ldots, S$, to be the photon counts received by the detector. It was observed when using confocal microscopes that "noise arising from photon is proportional to the square root of the signal intensity" (Wilson (1990)). If the photon counts c_j, $j = 1, \ldots, S$, are assumed to be independent then an idealised model can be that the c_j follow a Poisson distribution. The mean of this Poisson distribution is λ_i^* if pixel j is inside cell i. If j is a background pixel, then the mean of the Poisson distribution is zero, that is, the photon counts from background regions are zero with probability one. The records Y are the recordable version of the photon counts c after adding a dark-ground value and rescaling it. Assuming linear transformation is used to rescale the counts, the resulting records Y_j, $j = 1, \ldots, S$ can be expressed as

$$Y_j = a + bc_j,$$

for some constants $a, b \in \mathbb{R}$. To define a distribution of a record Y_j , $j = 1, \ldots, S$ based on the above transformations we use a Gaussian approximation of the distribution of the photon counts. The expected value and the variance of the records Y are therefore given by

$$E(Y_j) = \begin{cases} a & \text{if pixel } j \text{ is considered as background} \\ a + b\lambda_i^* & \text{if the pixel } j \text{ is inside the cell } i, \end{cases}$$

$$\mathrm{Var}(Y_j) = \begin{cases} 0 & \text{if pixel } j \text{ is considered as background} \\ b^2\lambda_i^* & \text{if pixel } j \text{ is inside the } i \text{ th cell,} \end{cases}$$

respectively. This is when using an idealised model. However, as mentioned in the previous section, because of the attenuation and the scattering effects of the light, the background signals are greater than zero. This is confirmed by the data and observed

variations within the background record values. To reflect the above ideas and using data analysis of the observed records we conclude the following:

- The value of the record $Y_j, j \in \{1, \ldots, S\}$, depends on the medium that pixel j falls in, whether it is a background region or a cell region.
- Once the physical structure of the specimen is determined, the records act independently, i.e., $Y_j|X\ j = 1, \ldots, S$ are independent. This assumption is supported by the space invariant property of confocal microscopy.
- The mean of the background region, τ_0 can be assumed to be fixed. $\tau_0 > 0$.
- There are differences in the values of the records in different cells which depend on the reflective index of the cells. Therefore, each cell i exhibits a mean $\tau_i = \tau_0 + \lambda_i$ where $\lambda_i = b\lambda_i^*$ can be considered as the added intensity level of the pixels lying inside the cell i.
- There is variation within the background records, ϕ_0^2 which can be considered as the variance of the white noise that is assumed to have a mean of zero.

Therefore based on the above we may assume that the experimental data $Y_j, j = 1, \ldots, S$ are normally distributed with mean

$$\mu_j = \begin{cases} \tau_0 & \text{if pixel } j \text{ is considered as background} \\ \tau_i & \text{if the pixel } j \text{ is inside the cell } i, \end{cases}$$

and variance

$$\sigma_j^2 = \begin{cases} \phi_0^2 & \text{if pixel } j \text{ is considered as background} \\ \phi_i^2 = \phi_0^2 + c^*(\tau_i - \tau_0) & \text{if pixel } j \text{ is inside the } i \text{ th cell,} \end{cases}$$

for some constants τ_0, ϕ_0^2 and c^*. The likelihood function of Y given $X = x$ is

$$\mathcal{L}(y|x) = \prod_{j=1}^{S} \frac{1}{\sqrt{2\pi}\sigma_j} \exp\{-\frac{1}{2\sigma_j^2}(y_j - \mu_j)^2\}. \tag{2.1}$$

2.2 Prior image model

2.2.1 Marked Point Process models

Our objective for this study is to make inference about the number of objects and their characteristics. Thus we require the prior model to handle certain geometric features of the objects such as their shapes, locations and their orientations. Baddeley and Van Lieshout (1993) have approached this problem by modelling objects as marked points distributed according to a specified point process. They represent each object in the scene as a pair (l, m), where l fixes the location of the object and the mark m contains all the required information to identify the object. Before we proceed with this approach we shall first define our object and its object space.

The class **U** of possible objects is an arbitrary set (called the object space). For example, the object space may be all different polygons on $\mathbb{R}^k$, where $k \in \mathcal{N}^+$ is the dimension of the space containing the object under interest. Or **U** could be restricted to a specified type of polygon. However, it needs not be a class of subsets of $\mathbb{R}^k$ if the specification includes properties like colours or surface texture.

Each object, X_i, will be modelled by a point l, representing its centre location within a specified window L, and a mark m which is a vector of variables, providing information about the object. The marks are required to lie in a specified space M. For example, in our 2D application, we approximate our objects by ellipses so l is the Cartesian coordinate of the centre of the ellipse and the marks m would be the semi-axes, angle of rotation and signal intensity (a more detailed descriptions will be given in chapter (3)).

An object configuration is simply a finite unordered set of objects $X = \{X_1, \ldots, X_N\}$,

$X_i \in \mathbf{U}, i = 1, \ldots, N$, where N is a random variable denoting the number of objects in L, and may be equal to zero.

To define the prior model of an object configuration, it is convenient to start with the ordered set of objects $(X_1, \ldots, X_N)$, and then to modify the results accordingly for the unordered set $X = \{X_1, \ldots, X_N\}$. Defining a distribution of $X = (X_1, \ldots, X_n)$ in $\mathbb{E} = \cup_n \mathbf{U}^n$, the space of all configurations X, is not easy and the distribution has to be built by reference to other processes.

2.2.2 The reference distribution

In the reference model, the location l of each object X_i is modelled by a Poisson process with intensity ρ, where we take ρ to be Lebesgue measure on L. Under the usual Poisson point process the distribution of the total number of points in L is

$$P(N = n) = \frac{\exp(-\rho(L))}{n!}\rho(L)^n.$$

Our reference model is an expansion of this point process model to a marked point process. The object space $\mathbf{U}$ is a product of two independent spaces L and M, $\mathbf{U} = L \times M$. Let ν define a probability measure on M such that $\nu(M) = 1$ and define the measure $\mu = \rho \otimes \nu$ on $\mathbf{U}$, combining ν with the above Poisson model for object locations. The mean number of objects in the set $\mathbf{U}$ is then $\mu(\mathbf{U})$, and for a set $A \subset \mathbf{U}$ it is $\mu(A)$. So we can write the probability of having a total of $N = n$ objects in $\mathbf{U}$ as

$$P(N = n) = \frac{\exp(-\mu(\mathbf{U}))}{n!}\mu(\mathbf{U})^n.$$

Given that exactly n objects are present, these objects X_i, $i = 1, \ldots, n$, are independent. Each X_i takes a value in $\mathbf{U}$ with probability measure $\mu(dx_i)/\mu(\mathbf{U})$. Hence, conditionally on $N = n$, $(X_1, \ldots, X_n)$ has measure

$$\frac{\mu(dx_1)}{\mu(\mathbf{U})} \cdots \frac{\mu(dx_n)}{\mu(\mathbf{U})}.$$

The reference measure, say $\Gamma(x,n)$, is then defined as the measure of the above reference process on $\mathbb{E} = \cup_n \mathbf{U}^n$. This is treating the $X_i, i = 1,\ldots,n$ as an ordered set, but it implies a distribution for the unordered objects, $\{X_1,\ldots,X_n\}$ on $\mathbb{E}^* = \cup_n \mathcal{U}^n$, the space of all configurations of $\{X_1,\ldots,X_n\}$ with respect to a reference measure, say $\Gamma^*(x,n)$. Here $\mathcal{U}^n$ is the space of the n unordered objects.

2.2.3 The model process

In order to construct a spatial process that exhibits interaction between neighbouring objects, we need to specify an absolutely continuous probability density $\pi(x,n)$, of the new process with respect to the reference process. This density should be a measurable and integrable function with respect to the reference measure $\Gamma(x,n)$ that maps $\mathbb{E} = \cup_n \mathbf{U}^n$ to $[0,\infty)$. In an intuitive sense $\pi(x,n)$ indicates how much more likely is the configuration of an event $X = x$ under our model than under the unrestricted Poisson process with an independent mark for each point.

Under the density $\pi(x,n)$ the probability of having n objects present is

$$P(N=n) = \frac{\exp(-\mu(\mathbf{U}))}{n!} \int_U \ldots \int_U \pi(x,n)\mu(dx_1)\ldots\mu(dx_n). \tag{2.2}$$

Given that there are exactly n objects, the joint probability density of the ordered set of objects $(X_1,\ldots,X_n)$ objects is

$$\pi(x|n) = z_n\, \pi(x,n), \tag{2.3}$$

with respect to the measure

$$\frac{\mu(dx_1)}{\mu(\mathbf{U})} \cdots \frac{\mu(dx_n)}{\mu(\mathbf{U})}.$$

Here z_n is the normalising constant and is equal to

$$z_n = [\mu(\mathbf{U})^{-n} \int_U \ldots \int_U \pi(x,n)\mu(dx_1)\ldots\mu(dx_n)]^{-1}.$$

Substituting into equation (2.2) we get

$$z_n = [\frac{\mu(\mathbf{U})^n \exp(-\mu(\mathbf{U}))}{P(N=n)n!}].$$

The density $\pi(x,n)$ can be used to exhibit *repulsion* or *inhibition* between objects. A widely used class for *pairwise interaction models* is

$$\pi(x,n) = \alpha\beta^{n(x)} \prod_{x_i \sim x_j} g(x_i, x_j), \tag{2.4}$$

where α is the normalising constant and $\beta > 0$ reflects the intensity of the process. We define the neighbourhood relation $\sim$ to be a symmetric, reflexive binary relation on $\mathbf{U}$, such that two objects are neighbours if their intersection is a non-empty set. Thus,

$$x_i \sim x_j \Leftrightarrow \Re(x_i) \cap \Re(x_j) \neq \phi$$

where $\Re(x)$ represents the image region occupied by the object x in L. $\Re(x)$ is determined by the l and m of x. The interaction function $g(\cdot,\cdot)$ is an integrable, non-negative bounded function that maps $\mathbf{U} \times \mathbf{U} \longrightarrow [0,\infty)$ which is specified by reference to the relation $\sim$. The product is over all distinct pairs of neighbouring objects $x_i \sim x_j$ with i < j. The object x may lie partially or entirely in L. A special case of (2.4) is the *hard core interaction model* which forbids objects to overlap, i.e., the interaction function g is zero if objects overlap and one otherwise, so the pairwise interaction model of $(X_1, \ldots X_n)$ is defined as

$$\pi(x,n) \propto \beta^{n(x)} \prod_{\text{all pairs } (x_i, x_j)} I[\text{no overlap between the objects } x_i \text{ and } x_j]. \tag{2.5}$$

In our study the objects represent cells. The cells are not permitted to overlap, so we will use the hard core interaction. For the unordered set we shall define the same model

as in (2.5) but with respect to $\Gamma^*(x,n)$ on $\mathbb{E}^* = \cup_n \mathcal{U}^n$.

For more extensive theoretical accounts on modelling and simulating spatial processes the reader can refer to Ripley (1981) and (1988), Diggle (1983), Baddeley and Van Lieshout (1993), Geyer and Moller (1994) and Van Lieshout (1995).

2.3 The posterior image distribution

Our posterior distribution $\pi(\cdot)$ is computed using *Bayes* theorem,

$$\pi(x,n|y) = \frac{\mathcal{L}(y|x,n)\pi(x,n)}{\int_{(x',n')\subset\mathbb{E}^*} \mathcal{L}(y|x',n')\pi(x',n')d\Gamma^*(x',n')}.$$

It will be used to make inference about the underlying scene X given the data $Y = y$. Since we cannot find the normalising constant we have to use MCMC so it suffices there just to have

$$\pi(x,n|y) \propto \mathcal{L}(y|x,n)\pi(x,n) \tag{2.6}$$

Chapter 3

Tailoring the model for the 2D problem

In this chapter we shall deal with a 2D image frame and shall specify a statistical model for the 2D space; the 3D case will be handled separately in chapter (6).

3.1 Likelihood model

We consider an optical section from the stack of 2D sections which forms the 3D image obtained from the fluorescent confocal microscope. The experimental data consist of an image $Y = (Y_j; j = 1, \ldots, S)$ where S indexes a 184×768 grid (see figure $(3-1)$). It should be noticed that in all the optical section resolutions in this thesis the $x-$axis is the vertical axis and accordingly the $y-$direction is the horizontal one. We assume that the $Y_j, j = 1, \ldots, S$, are conditionally independent given $X = x$, and normally distributed. The mean of this distribution will be given by

$$\mu_j = \begin{cases} \tau_0 & \text{if pixel } j \text{ is considered as background} \\ \tau_i & \text{if the pixel } j \text{ is inside the cell } i \end{cases} \tag{3.1}$$

Figure 3-1: A 2D optical section collected by confocal fluorescence microscopy for an area of cartilage growth.

and the variance of the record Y_j is

$$\sigma_j^2 = \begin{cases} \phi_0^2 & \text{if pixel } j \text{ is considered as background} \\ \phi_i^2 = \phi_0^2 + c^*(\tau_i - \tau_0) & \text{if pixel } j \text{ is inside the } i \text{ th cell.} \end{cases} \tag{3.2}$$

Here τ_0, ϕ_i^2, c^* are constants which we have determined by analysing some training data. We assign a value of 38.0 to τ_0, the mean of the background region (see figure $(3-2)$). The variance of background pixels ϕ_0^2 was estimated using a sample of records and set to 46.50. To determine the value of c^* we plotted sample variances σ^{2*} of the cells against their sample means μ^* (see figure $(3-3)$). There are several models which could explain the relation between the mean and the variance of the cells. For example, using the *least squares*, we fit the following linear regression model

$$\sigma_i^{2*} = -488.9 + 7.02\mu_i^*.$$

The cubic regression fit for these data implies

$$\sigma_i^{2*} = 1006.31 - 36.75\mu_i^* + 0.376\mu_i^{2*} - 0.001\mu_i^{3*}.$$

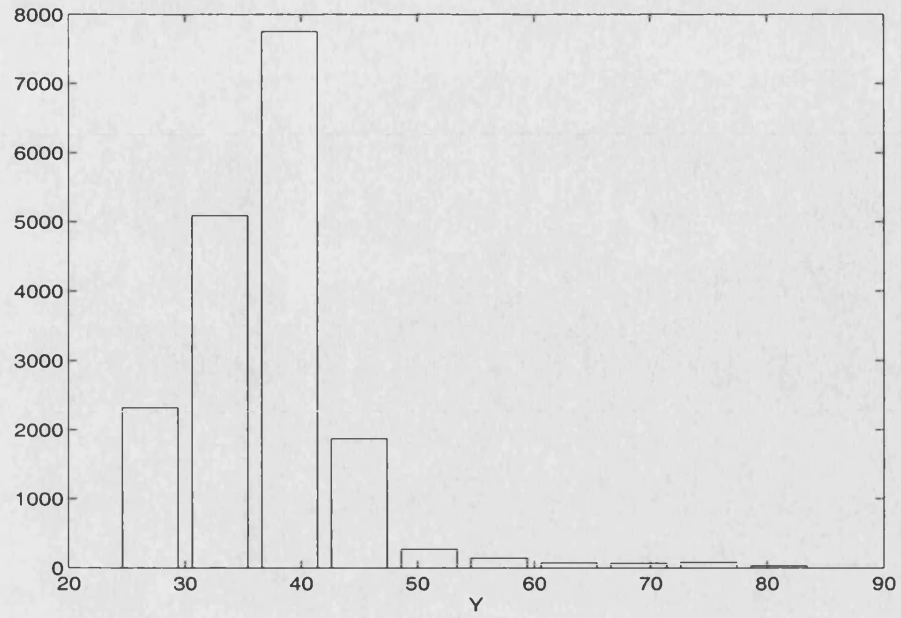

Figure 3-2: A histogram of a sample of pixels in a background region.

A general form for linear relations to fit is given by

$$\sigma_i^{2*} = \mathbf{a}^* + \mathbf{b}^* \mu_i^*, \tag{3.3}$$

where $\mathbf{a}^*$ and $\mathbf{b}^*$ are constants with $\mathbf{b}^* > 0$. Plots of the above relations for three different pairs of values for $(\mathbf{a}^*, \mathbf{b}^*)$=(8.5,1),(−67,3) and (−125,4.5) are displayed in figure $(3-3)$.

It is clear that there are high scatters about the above models. This is due to the variation in the records: the variation of the Poisson distribution, blurring effects, attenuation effects and other sources of variation which will be discussed later in more detail. Meanwhile we shall choose to work with a simple linear one, preliminarily setting $\mathbf{a}^* = 8.5$ and $\mathbf{b}^* = 1$ which implies $c^* = 1$ in equation (3.2). That is because we believe that the differences between these models are not crucial as long as they do not affect the reconstruction of the data negatively. None of these fitted linear models describes the relation between the sample means and sample variances accurately. Some are better

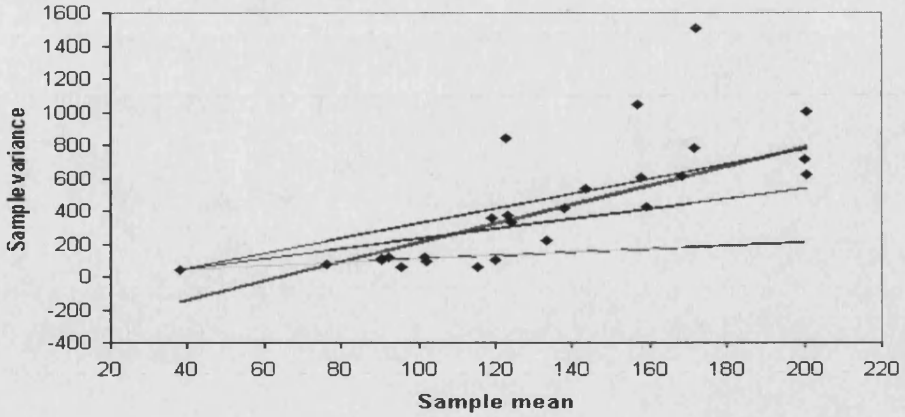

Figure 3-3: A plot of the sample variances of the cells against their samples means. The lines represent the possible choices for the linear relations between the mean and the variance. The bold line is the LSE linear regression line.

estimators for the small sample means and some are better for the large sample means. The chosen model is more precise in explaining the relation for low level intensity cells where there is the greatest risk of classifying the cell area as background. For the high intensity cells this model will eventually produce the same results as the other models concerning their reconstruction.

3.2 Object description

Although the cells under study do not have completely regular shapes (see figure (3−4)), a good approximation for describing the cells is as ellipses. The standard form of the equation of the surface of an ellipse with centre at (c_x, c_y) and horizontal and vertical semi-axes is given by

$$\frac{(x-c_x)^2}{a^2} + \frac{(y-c_y)^2}{b^2} = 1$$

where a is the first semi-axis of the ellipse representing its width, and b is the second semi-axis representing its length. The semi-axis a is parallel to the $x-$axis while b is

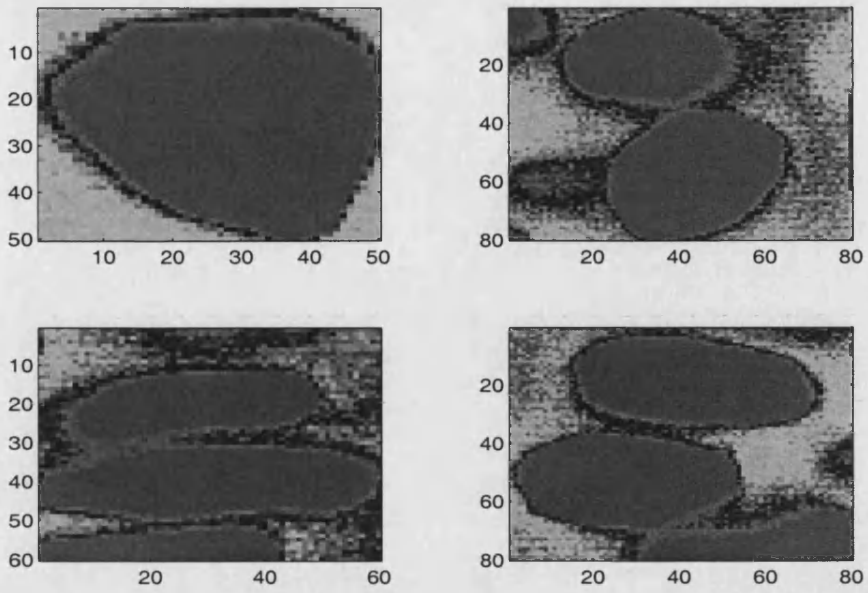

Figure 3-4: Images for some of the cells in the grid using the input records.

parallel to the $y-$axis. We shall consider ellipses which lie in different orientations as seen in figure $(3-1)$. To allow cells to take different orientations in the scene we shall consider rotations of the standard form of ellipses around their centre points. In the rotation of axes (orientation) process we introduce a new position of an ellipse system by keeping its centre point (c_x, c_y) fixed and rotating the semi-axes a, b about the centre to another position. We shall define θ to be the angle of rotation (see figure $(3-5)$) . The equation of the ellipse after rotating through an angle θ is now

$$\frac{((x-c_x)\cos(\theta)+(y-c_y)\sin(\theta))^2}{a^2}+\frac{(-(x-c_x)\sin(\theta)+(y-c_y)\cos(\theta))^2}{b^2}=1$$

Each ellipse also exhibits an intensity level λ which can be considered as the added intensity levels of the pixels lying inside that cell. Referring to equation (3.1), suppose a pixel j lies inside an object i then, the expected value of the record of pixel j is the

sum of intensity level of cell i and the intensity mean of the background area, i.e.,

$$\tau_j = \lambda_i + \tau_0.$$

3.3 The prior model

To construct the prior distribution for the 2D case we shall follow the same steps defined in section (2.2). We shall start with identifying the object space $\mathbf{U} = L \times M$ and defining a measure μ on $\mathbf{U}$. This is followed by defining the space of all the configuration of X of the model, $\mathbb{E}$ and the reference measure $\Gamma(X, N)$ in $\mathbb{E}$. In $\mathbb{E}$ we shall mark out the model process as a density $\pi(x, n)$ with respect to $\Gamma(X, N)$.

Each elliptical cell is then specified by a marked point $x = (l, m)$, the point l being the location of the centre of the ellipse. The location l is a continuous variable required to lie in the window L, where L is defined to be the continuous space covered by an arbitrary fixed shape and fixed size window. Initially we define our window L to be the continuous space covering the rectangular lattice $[1 : S_x, 1 : S_y]$. Practically l is a 2D point and consists of an ordered pair of coordinates, $l \in L = \{(c_x, c_y) : c_x \in [0, L_x = 184), c_y \in [0, L_y = 768)\}$. If we are interested in the cells with centres outside the window but with part of the ellipse area inside the window then we can enlarge the existing window. We set the new window L' to be $[-o, L_x + o) \times [-o, L_y + o)$ where o is an arbitrary non negative real value constant. By defining the new window L' we can consider the cells with centres in L as well as the cells with centres lying in $L'|L$ with part of the cells lying in the original window L. The values of the records for the pixels lying outside the original window L and inside the extended window L' can be treated as missing values. The locations of the objects are assumed to be uniformly distributed over the window L or L'.

The mark m is a vector of four independent variables (a, b, θ, λ) assigning values for the first semi-axis, second semi-axis, angle of rotation and the intensity of the cell

respectively. We shall assume that the lengths of the axes are bounded by some pre-determined range $M_{axis} = (\min_{axis}, \max_{axis})$, and that the intensity λ lies within the range $(\min_\lambda, \max_\lambda)$. The general state space of the mark m is then the product of the spaces of its components, $M = (\min_{axis}, \max_{axis}) \times (\min_{axis}, \max_{axis}) \times (0, \pi) \times (\min_\lambda, \max_\lambda)$. The location l takes a value on $L \subset \mathbb{R}^2$ and mark m takes a value in $M \subset \mathbb{R}^4$, so a single object $(c_x, c_y, a, b, \theta, \lambda)$ has to lie in $\mathbf{U} = L \times M \subset \mathbb{R}^6$. So we can define the object space $\mathbf{U}$ to be the set that contains the representations of ellipses defined in $\mathbb{R}^2$ with centres inside the window L and different semi-axes, angles of orientation and intensities which are required to be in a specified range. It can be noted that each ellipse has two representations $(l, a, b, \theta, \lambda)$ and $(l, b, a, \hat{\theta}, \lambda)$, where $|\theta - \hat{\theta}| = \pi/2$.

To create the reference process, as defined in section (2.2.2), we suppose the collection of points is a Poisson process on L with intensity one. We denote the area of L by $\mathcal{A}_L$. For each point, we define independently a mark $m \in M$ from a probability measure ν on M. As we defined $M \subset \mathbb{R}^4$, then we can express the $\sigma-$finite measure ν as a density with respect to Lebesgue measure. In order to model the variability of the cells we specify suitable distributions for the mark variables that reflect what we expect to observe in the data. The semi-axes a and b are assumed to be independent, each is defined in M_{axis} and their density is proportional to that of a bivariate normal with mean (μ_a, μ_b) and covariance matrix

$$\Sigma = \begin{bmatrix} \gamma^2 & 0 \\ 0 & \gamma^2 \end{bmatrix}.$$

We shall have $\mu_a > \mu_b$ but we do not insist that $a > b$, so a and b are not necessarily the major and the minor semi-axes, respectively. For θ we have defined a probability density

$$\pi_\theta(\theta) = \frac{|\cos(\theta)| + \frac{1}{\pi}}{3}, \; 0 \leq \theta \leq \pi.$$

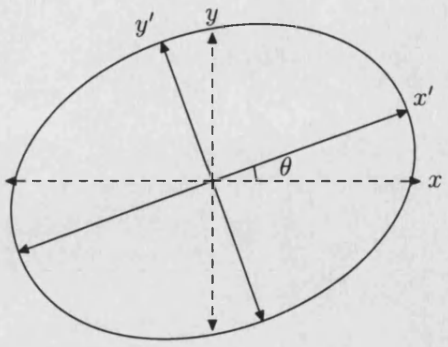

Figure 3-5: The angle of rotation θ of the $xy-$axis around the centre of the ellipse.

Defining this density for θ (see figure $(3-6)$)will generate values close to 0 or π more frequently than other values on $(0,\pi)$, reflecting the observed positions of the cells (see figure $(3-1)$). Finally for the intensity rate λ we assume that it is uniformly distributed over a specific range $M_\lambda = (\min_\lambda, \max_\lambda)$.

It follows that in the reference model discussed in section (2.2.3) we can define for a single object X_i a density

$$f_{X_i}(c_x, c_y, a, b, \theta, \lambda) = \frac{1}{\mathcal{A}_L}\frac{z}{2\pi\gamma^2}\exp(-\frac{1}{2\gamma^2}\{(a-\mu_a)^2+(b-\mu_b)^2\})$$
$$\frac{|\cos(\theta)|+\frac{1}{\pi}}{3}\frac{1}{\max_\lambda - \min_\lambda} I[(c_x,c_y)\in L,\ a,b\in M_{axis},\ \theta\in(0,\pi),\ \lambda\in M_\lambda], \qquad (3.4)$$

with respect to Lebesgue measure μ on $\mathbf{U}$. Here z is a normalising constant for (a,b). For the rest of this thesis, we shall write f_{X_i} without the indicator I of the variables involved in the mark or the location of X except for (a,b) as it implies a normalising constant.

Following the steps of section (2.2) for defining the prior model for an object configuration $X = \{X_1,\ldots,X_n\}$, we shall first define the prior distribution for the or-

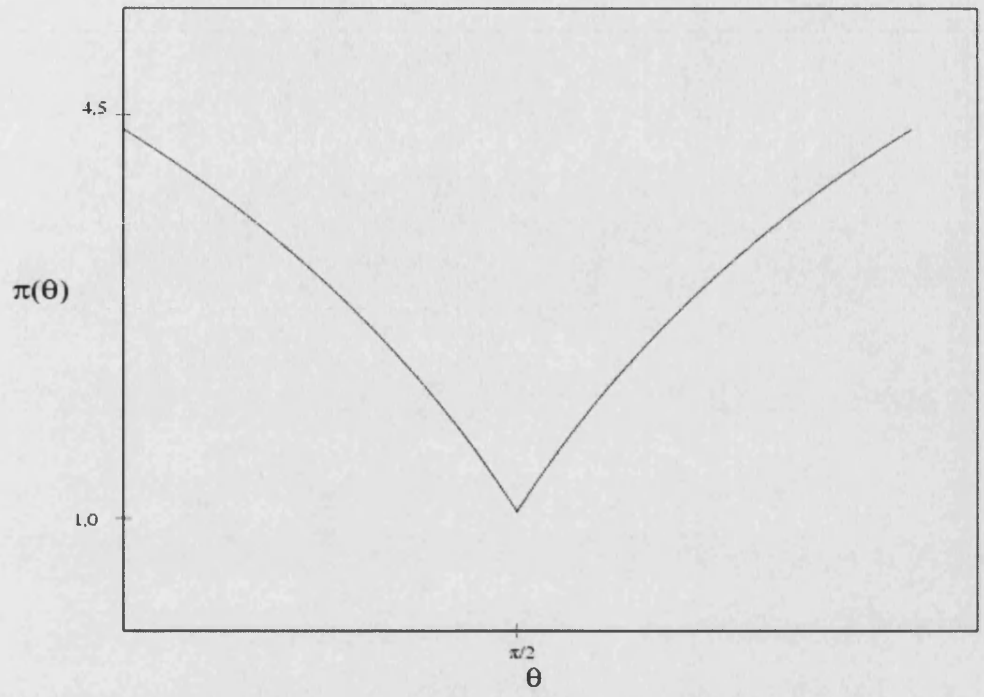

Figure 3-6: The distribution of the angle of rotation $\pi_\theta(\theta)$.

dered objects. Suppose that an object configuration is a finite list of ordered objects $X = (X_1, \ldots, X_n)$, $X_i \in \mathbf{U}$, $i = 1, \ldots, n$. The formulation of the states of the model is that for each configuration $X \in \mathbb{E} = \cup_n \mathbf{U}^n$ there are two components: an index N representing the total number of objects in the current configuration and a value from a set $\mathbf{U}^N \subset \mathbb{R}^{6N}$ that depends on the index. For example, for an empty configuration $N = 0, X = ()$, an empty list, the state space is $\mathbf{U}^0$. Combining the Poisson process of the cells with their associated mark distribution gives the reference measure $\Gamma(X, N)$.

To model the fact that the cells do not overlap we define a density $\pi(x, n)$ with respect to the reference measure $\Gamma(X, N)$. This density should exhibit interaction

between neighbouring objects. Thus we use the hard core model

$$\pi(x,n) = \alpha\beta^n I[\text{no overlap between the cells in the scene}], \tag{3.5}$$

where α is the normalising constant and the constant β reflects the mean number of objects expected under the density $\pi(x,n)$. The reference measure $\Gamma(X,N)$ in this problem is defined in $\mathbb{E} \subset \cup_{n=0}^{\infty}\mathbb{R}^{6n}$. It comprises two parts: the total number of objects N which will have a Poisson distribution with mean $\mathcal{A}_L$, and the objects themselves $X_1,\ldots,X_N$, each having a density

$$f_{X_i}(c_x,c_y,a,b,\theta,\lambda) = \pi_l(c_x,c_y)\pi_{a,b}(a,b)\pi_\theta(\theta)\pi_\lambda(\lambda), \tag{3.6}$$

with respect to μ on $\mathbf{U}$.

Later on, in implementing the Metropolis-Hastings algorithm, we shall need to do detailed calculations to find the acceptance probability of proposed moves. For these purposes it is convenient to express the posterior image distribution as a collection of separate parts, each of which has a density in a space of one particular dimension. With this in mind, we start by presenting the reference measure $\Gamma(X,N)$ as a collection of sub-measures $\Gamma_n(X)$, $n = 0,1,2,\ldots$. Each $\Gamma_n(X)$, $n = 0,1,2,\ldots$ is defined for values of x on a different part of the state space $\mathbb{E}$. For states with $N = n$ objects, the part of the reference measure $\Gamma_n(X)$ is confined to $x \in \mathbf{U}^n \subset \mathbb{R}^{6n}$ and here X has conditional probability measure $\mu(dx_1)/\mu(\mathbf{U})\ldots\mu(dx_n)/\mu(\mathbf{U})$, given $N = n$. This conditional distribution of $(X_1,\ldots,X_n)$ given $N = n$ has a density

$$\prod_{i=1}^{n} f_{X_i}(x_i). \tag{3.7}$$

with respect to Lebesgue measure on $\mathbb{R}^{6n}$, where f_{X_i} is as defined on equation (3.6).

In the part of space $\mathbb{E}$ where $N = n$, we can combine the probability that $N = n$ with the conditional density of X given $N = n$ to express the sub-probability measure

$\Gamma_n(X)$ as a sub-density with respect to Lebesgue measure on $\mathbb{R}^{6n}$, for this particular n,

$$\frac{e^{-\mathcal{A}_L}\mathcal{A}_L^n}{n!}\prod_{i=1}^{n} f_{X_i}(x_i),\ \ x_i \in \mathbf{U},\ \ i=1,\ldots,n. \tag{3.8}$$

Note that, this is not a conditional distribution of X given $N=n$ but rather a joint distribution of N and X restricted to $N=n$ objects. It follows that, the prior model in this part of the state space has a sub-density $\tilde{\mathcal{W}}_{X,N}(x,n)$ with respect to Lebesgue measure on $\mathbf{U}^n$ equal to

$$k\beta^n \frac{e^{-\mathcal{A}_L}\mathcal{A}_L^n}{n!}\prod_{i=1}^{n} f_{X_i}(x_i) I[\text{no overlap between the cells in the scene}], \tag{3.9}$$

where k is a normalising constant for the whole distribution and does not depend on n. As a result of the overlapping condition the objects $X_1,\ldots,X_n$ are not independent given $N=n$.

This is treating the list of objects $(x_1,\ldots,x_n)$ as an ordered set. Although it is convenient to work with the ordered list of objects, we really think of an image as an unordered set of objects $x=\{x_1,\ldots,x_n\}$. We, therefore, need to deduce the prior distribution for the unordered sets of objects. The distribution of the ordered lists $(X_1,\ldots,X_n)$ implies a distribution for the sets $\{X_1,\ldots,X_n\}$. The distribution of the unordered set will be defined in rather odd space, say $\mathcal{U}^n$. The space $\mathcal{U}^n$ of the unordered set $\{x_1,\ldots,x_n\}$ can be defined from that of the ordered $(x_1,\ldots,x_n)$ by choosing any one variable of the mark or the location of $x_i=(c_{x_i},c_{y_i},a_i,b_i,\theta_i,\lambda_i)$. Here, we choose the angle of rotation θ. Suppose we define the space of the unordered set of the n angles to be Δ_n then the space of the unordered set of the $n-$objects $\{x_1,\ldots,x_n\}$ is just the product of $\Delta_n \times \{L \times M_{axis} \times M_{axis} \times M_\lambda\}^n$. As all the variables involved in defining an object are continuous variables, there is zero probability of having two objects with equal values, hence, the ordered list of the n objects is unique. Let $(\theta_1,\ldots,\theta_n)$ denote the ordered list of θ_i , $i=1,\ldots,n$, of the corresponding objects. The $n-$dimensional region Δ_n for $\{\pi \geq \theta_1 > \theta_2 > \ldots > \theta_n \geq 0\}$, is an $n-$dimensional

simplex. A zero-dimensional simplex is a point a_0. A one-dimensional simplex is a segment $[a_0, a_1]$ and is defined in $\mathbb{R}^2$. A two-dimensional simplex is a triangle $[a_0, a_1, a_2]$ with 3-faces obtained by joining the vertices $a_i, i = 0, 1, 2$. The two-dimensional simplex is defined in $\mathbb{R}^3$. A three-dimensional simplex is a tetrahedron $[a_0, a_1, a_2, a_3]$. Hence, by induction an n-dimensional simplex $[a_0, \ldots, a_n]$ is defined and lies in an n-dimensional space $\mathbb{R}^n$. Thus $\mathcal{U}^n \subset \mathbb{R}^{6n}$, hence, we can express the sub-measure defined on $\mathcal{U}$ as a density for $\{X_1, \ldots, X_n\}$ with respect to a Lebesgue measure. This density, say $f_{\{X\}_N}$, is defined in terms of $f_{(X)_N}(X) = \prod_{i=1}^{N} f_{X_i}(x_i)$, the density for the ordered list of N objects. To illustrate: if $N = 1$ then $\mathcal{U} = \mathbf{U}$ and, hence, no change in the space nor the density $f_{(X)_1}(x)$ defined for this space $\mathbf{U}$ has occurred. For a two object ordered set we have defined $f_{(X)_2}(x)$ with respect to the Lebesgue measure on $\mathbf{U}^2 = M_\theta \times M_\theta \times \{L \times M_{axis} \times M_{axis} \times M_\lambda\}^2$. Here $M_\theta \times M_\theta = M_\theta^2$ represent the area of rectangle with length $M_\theta = (0, \pi)$. Whereas for the unordered set of two objects the space is $\triangle_2 \times \{L \times M_{axis} \times M_{axis} \times M_\lambda\}^2$. The area of the $\triangle_2$ is $\frac{1}{2}M_\theta^2$. Each two points, (θ_1, θ_2) and (θ_2, θ_1), in $M_\theta \times M_\theta$ map into one point, $\{\theta_1, \theta_2\}$, in $\triangle_2$. Thus the density $f_{\{X\}_2}$ of the unordered set $\{x_1, x_2\}$ is equal to $2f_{(X)_2}(x)$. For the three objects set the volume of space of the ordered set is 6-times the volume of the unordered one. Therefore the density defined for $\mathcal{U}^3$ is 6 times the probability density for $\{X_1, X_2, X_3\}$, which is equal to $3!f_{(X)_3}$. By induction for $N = n$ objects the measure of the space of the ordered set of these objects is $n! \times \{$the measure of the space of the unordered set$\}$, i.e.,

$$Volume(\mathcal{U}^n) = \frac{Volume(\mathbf{U}^n)}{n!}.$$

The density of the unordered set of $\{X_1, \ldots, X_n\}$ is defined by considering the density of the ordered objects and multiplying it by a factor $n!$, corresponding to the number of equally likely permutations of $\{X_1, \ldots, X_n\}$. For the part of the state space $\mathbb{E}$ with $N = n$ objects, the unordered set of objects $x = \{x_1, \ldots, x_n\}$ in $\mathcal{U}^n$ has a sub-density $\mathcal{W}_{X,N}(x, n)$ with respect to Lebesgue measure equal to $n!\tilde{\mathcal{W}}_{\tilde{X},N}(\tilde{x}, n)$. Here $\tilde{x}$ denotes

an ordered list of $x_1, \ldots, x_n$. Substituting for $\tilde{\mathcal{W}}$ in equation (3.9), $\mathcal{W}_{X,N}(x,n)$ becomes

$$k\frac{e^{-\mathcal{A}_L}\beta^n}{(\max_\lambda - \min_\lambda)^n}\{\prod_{i=1}^{n}\frac{|\cos(\theta_i)|+\frac{1}{\pi}}{3}\frac{z}{2\pi\gamma^2}\exp(-\frac{1}{2\gamma^2}(a_i-\mu_a)^2+(b_i-\mu_b)^2)$$
$$I_{[a_i\in M_{axis},b_i\in M_{axis}]}\}I[\text{no overlap between the cells in the scene}]. \quad (3.10)$$

3.4 The reversible jump algorithm

In implementing our MCMC simulations, we shall define the following move types: adding an object; shifting, resizing, rotating or changing the intensity of a cell; deleting an existing cell, merging two objects, or splitting an existing object. Since the moves: adding (*birth*) and deleting a cell (*death*) or merging two objects and splitting an object, induce changes in the dimension of the parameter vector, we need MCMC algorithms which are able to simulate from variable dimension posterior models. Standard MCMC methods cannot do this but Green (1995) has derived the *reversible jump algorithm* which can be used here.

3.4.1 Introduction

Suppose it is desired to sample a value of X from a probability measure Π on a sample space Ω, where parts of Ω are of different dimensions. The reversible jump algorithm is a MCMC sampler designed to handle these changes in dimension. It is an extension of the Metropolis-Hastings method to cases in which the state space is the countable union of subspaces of differing fixed dimensions. That is, the reversible jump algorithm is essentially Metropolis-Hastings algorithm with measures rather than densities. Green was not the first one to present an algorithm for these type of problems. Geyer and Moller (1994) have developed an MCMC sampler for *simulating spatial point processes* as an alternative for *spatial birth-and-death processes* (Preston (1977) and Moller (1989)) which converges to the Gibbsian point process. Geyer and Moller derived likelihood inference procedures for point pattern based on point process. Grenander and

Miller (1994) proposed *jump-diffusion* processes for Bayesian computation in certain computer vision problems. In this algorithm they limit the range of jump transitions between parameter subspaces such that the proposals are generated from the prior. The above algorithms: spatial birth-and-death processes, simulating spatial point processes and jump-diffusion algorithm are very close to the reversible jump algorithm and can be considered as special cases of reversible jump algorithm (see Green (1994) and (1995), Geyer (1996) and Cappe *et al.* (2001)). In Green's algorithm, the state space is traversed using a countable set of move types some of which jump between subspaces of different dimension. In the image problem, such moves are birth and death of a cell or split of a cell into two and merging of two cells. All move types, or pairs of move types in the cases of birth and death or merge and split, attain detailed balance individually.

3.4.2 The general case

In state x, the reversible jump algorithm chooses a move of type v with probability $j_v(x)$, $v \in \{1, \ldots, V\}$. Conditional on the choice of v, x' is proposed according to rules defined for this move type. Let $K_v(x, dx')$ denote the conditional kernel used to generate x' given that a move type v is chosen, then the joint kernel, $Q_v(x, dx') = j_v(x)K_v(x, dx')$, is a sub-probability measure on x'. This proposal kernel $Q_v(x, dx')$ depends on the starting state x and the move type v. In general, not all the move types can propose x' from x, hence, for these moves we set $Q_v(x, dx') = 0$. For each move type v there is a reverse move of type v' that takes the state x' back to x with sub-probability measure $Q_{v'}(x', dx)$ specifying the joint distribution of v' and x. Green shows that setting the acceptance probability for the move of type v from x to x' to be

$$\alpha_v(x, x') = \min\left\{1, \frac{\Pi(dx')Q_{v'}(x', dx)}{\Pi(dx)Q_v(x, dx')}\right\} \tag{3.11}$$

achieves detailed balance. Here α_v is a measurable function that maps $\Omega \times \Omega \longrightarrow [0,1]$. For the reverse move v' that takes the state x' back to x the acceptance probability is

$$\alpha_{v'}(x', x) = \min\left\{1, \frac{\Pi(dx)Q_v(x, dx')}{\Pi(dx')Q_{v'}(x', dx)}\right\}.$$

The terms appearing in the ratio of the right hand side of equation (3.11) are very general and further detail is needed to explain their meaning in particular cases. Equation (3.11) is best understood by comparison with the corresponding Metropolis-Hastings update in equation (1.6), where the unnormalised density π is replaced by an unnormalised measure Π on the state space Ω and the proposal density $q(x, x')$ is now replaced by proposal kernel $Q_v(x, dx')$. Tierney (1998) defined the numerator and the denominator of equation (3.11) as measures on some state space and specified a common symmetric measure, say ξ, on $\Omega \times \Omega$ used for that general state space. He described the necessary conditions for the measure ξ to dominate $\Pi(dx)Q_v(x, dx')$ and $\Pi(dx')Q_{v'}(x', dx)$ so that there are a Radon-Nikodym derivates

$$h_v(x, x') = \frac{\Pi(dx)Q_v(x, dx')}{\xi(dx, dx')}, \text{ and } h_{v'}(x', x) = \frac{\Pi(dx')Q_{v'}(x', dx)}{\xi(dx', dx)}$$

which replace $\pi(x)q(x, x')$ and $\pi(x')q(x', x)$, respectively, in the ordinary Metropolis-Hastings equation. Tierney also defined the conditions on the acceptance probability function α_v for the resulting transition kernel and the invariant distribution to satisfy detailed balance condition. We shall give some illustrative examples in the rest of this section that show what these notations can mean in certain cases.

If x and x' have the same dimension, say $x, x' \in \tilde{\Omega}$. That is, when move type v preserves the dimension of the state, then v and v' are the same move type and the ratio term in the above equation reduces to a simple density ratio. Hence, in this case the ratio

$$\frac{\Pi(dx')Q_{v'}(x', dx)}{\Pi(dx)Q_v(x, dx')}, \tag{3.12}$$

can be implemented as the product of two terms. The first, $\pi(x')/\pi(x)$, is the ratio of the density of Π with respect to the same $\sigma-$finite measure on $\tilde{\Omega}$ evaluated at x' and at x respectively. The second term corresponds to the the kernels' ratio $Q_{v'}(x', dx)/Q_v(x, dx')$, and can be written as the ratio of the proposal densities, $q_v(x', x)$ and $q_v(x, x')$, with respect to a common dominating $\sigma-$ finite measure. To illustrate, we use a simple example for an object recognition problems where interest is only in the location of each object, $x_i = (l_i)$ which takes a value in $\mathbb{R}$. Then an object configuration is $(x_1, \ldots, x_n) \in \mathbb{R}^n$. The target measure Π is then defined on $\Omega = \emptyset \cup \mathbb{R} \cup \mathbb{R}^2 \cup \ldots$. Suppose we define a shift move in which we propose to update a randomly picked object x_i to x_i' using the proposed kernel q_{shift}. Suppose that we are using a random update sampler, i.e., a sampler iteration is completed by this single step. Let $x' = (x_1, \ldots, x_i', \ldots, x_n)$ be the resulting configuration after replacing x_i by x_i'. The dimension of configurations x and x' is the same and is equal to n. As $x \in \mathbb{R}^n$ and $x' \in \mathbb{R}^n$, they both have a sub-probability density $\pi_{N=n}$ with respect to Lebesgue measure on the space $\mathbb{R}^n$. The ratio $\Pi(dx')/\Pi(dx)$ is the ratio of the sub-probability density π_n evaluated at x' over that sub-density evaluated at x, that is $\pi_n(x')/\pi_n(x)$. The kernel $Q_v(x, dx')$ in this case can be written as $q_{shift}(x_i, x_i')$ which is a density but with respect to a Lebesgue measure on $\mathbb{R}$. Switching to state x from state x' is done by considering the same move type. We propose to move from x' to x using the same proposing kernel h_{shift}, hence, $v = v'$ and the ratio $Q_{v'}(x', dx)/Q_v(x, dx')$ defined in equation (3.12) can be written here as $q_{shift}(x_i', x_i)/q_{shift}(x_i, x_i')$ which is defined in $\mathbb{R}$.

The situation is more complicated when x and x' are from two subspaces with different dimensions. If we specify a move type *birth* where we propose to add a new object x_{n+1}, then the starting configuration $x = (x_1, \ldots, x_n)$ is in $\mathbb{R}^n$ while the proposed configuration $x' = (x_1, \ldots, x_n, x_{n+1})$ is in $\mathbb{R}^{n+1}$. So we cannot use the same interpretation for the acceptance rate as we did in the shift move. As the target distribution Π is assumed to be well defined in these subspaces, then $\Pi(dx)Q_v(x, dx')$ and $\Pi(dx')Q_{v'}(x', dx)$ define measures on the product of the general states $\Omega \times \Omega$. The

ratio term

$$R_v = \frac{\Pi(dx')Q_{v'}(x', dx)}{\Pi(dx)Q_v(x, dx')}, \tag{3.13}$$

can be interpreted as a ratio of the densities of these measures with respect to some common underlying measure on $\Omega \times \Omega$. The measure exists providing that $Q_v(x, dx')$ and $Q_{v'}(x', dx)$ satisfy a *dimension matching* condition. This condition requires the set of possible pairs (x, x') for a move of type v should have the same *dimension* as the set of possible pairs (x', x) for a move of type v'. Essentially, one may consider the pair $(x, x') \in \mathbb{R}^n \times \mathbb{R}^{n+1} = \mathbb{R}^{2n+1}$ and $(x', x) \in \mathbb{R}^{n+1} \times \mathbb{R}^n = \mathbb{R}^{2n+1}$, however, the constraints on the moves which are possible imply that ΠQ_v can be expressed as a density with respect to Lebesgue measure on $\mathbb{R}^{n+1}$ and so does $\Pi Q_{v'}$. Hence, the sets of possible pairs (x, x') and (x', x) are meant to be in $\mathbb{R}^{n+1}$. More explanation on this considering the general treatment of the reversible jump algorithm is given in the next section.

3.4.3 Switching between two subspaces

Although moves between spaces of different dimensions may sound complicated, especially if we are dealing with measures on a general multi-dimensional state space, this can be done simply with respect to Lebesgue measure. Consider a move type v taking the current state $x \in \mathbb{R}^{n_1} \subset \Omega$ to state x' in a higher dimension space $x' \in \mathbb{R}^{n_2} \subset \Omega$, where $n_2 - n_1 = d > 0$. The move v is to be chosen with probability $j_v(x)$. Suppose we can make this move by generating a vector of continuous random variables u of size d from a proper density q with respect to Lebesgue measure on $\mathbb{R}^d$, independently of x, and setting x' to be some deterministic function f of x and u, $x' = f(x, u)$. It is convenient to assume that each set of values of (x, u) will give a different value for x'. We can express the measure of the transition kernel $Q_v(x, dx')$ as a sub-density

$$j_v(x)q(u)\left|\frac{d(x,u)}{dx'}\right|,$$

with total mass $j_v(x)$ with respect to Lebesgue measure on a d dimensional subspace of $\mathbb{R}^{n_2}$. That is

$$\begin{aligned}\int_{x' \in \Omega} Q_v(x, dx') &= j_v(x) \int_{x' \in \Omega} q(u) \left| \frac{d(x,u)}{dx'} \right| dx' \\ &= j_v(x) \times 1 \\ &= j_v(x).\end{aligned}$$

The *Jacobian* term arises because we are taking the ratio of two measures each defined in a different space. The numerator is defined for the variables (x, u) and the denominator is defined for the transformed variable x' which is a function of the former variables.

Now consider a move type v' from a value of x' in $\mathbb{R}^{n_2}$ to $x \in \mathbb{R}^{n_1}$. For the process to be reversible, the move from x' should be to a value x such that $x' = f(x, u)$ for some u and, by our previous assumption, there is a unique pair (x, u) for which $x' = f(x, u)$. This needs the moves v and v' to define a bijection between the set of values of (x, u) defined in $\mathbb{R}^{n_1} \times \mathbb{R}^d$ and the set of values of x' which are defined on $\mathbb{R}^{n_2}$. We have achieved this by choosing u of length $d = n_2 - n_1$ and defining in the inverse move v' a map of (x', x) satisfying the relation $x' = f(x, u)$. Let $j_{v'}(x')$ be be the probability of choosing move v' starting from state x'. Suppose Π has sub-densities π_{n_1} and π_{n_2} on $\mathbb{R}^{n_1}$ and $\mathbb{R}^{n_2}$ respectively and there is no action to be taken except to determine x directly from x' and set $x = x(x')$, then the joint density of x' and x, $\Pi(dx')Q_{v'}(x', dx)$ in the numerator of equation (3.13) is $\pi_{n_2}(x')j_{v'}(x')I[x = x(x')]$.

The acceptance probability for a move from x to x' where $x = x(x')$ becomes

$$\alpha_v(x, x') = \min\{1, R_v\},$$

where

$$R_v = \frac{\pi_{n_2}(x')j_{v'}(x')}{\pi_{n_1}(x)j_v(x)q(u)} \left| \frac{dx'}{d(u,x)} \right|, \tag{3.14}$$

and that for the reverse move v' that takes the state from x' to x is,

$$\alpha_{v'}(x', x) = \min\{1, R_v^{-1}\}.$$

To see that this construction gives a Markov chain with a stationary distribution Π, we shall demonstrate detailed balance by showing that , for any $A \subset \mathbb{R}^{n_1}$ and $B \subset \mathbb{R}^{n_2}$,

$$\int_A \Pi(dx) \int_B Q_v(x, dx')\alpha_v(x, x') = \int_B \Pi(dx') \int_A Q_{v'}(x', dx)\alpha_{v'}(x', x). \tag{3.15}$$

Suppose without loss of generality that starting at B we will end in A under move v' with probability one, then the above condition can be written as

$$\int_{x\in A} \pi_{n_1}(x)j_v(x)P\{\text{Accept an } X' \in B|x, v\}dx = \int_{x'\in B} \pi_{n_2}(x')j_{v'}(x')\alpha_{v'}(x', x)dx'. \tag{3.16}$$

Using the random variable u to switch from x to x', the left hand side of equation (3.16) becomes

$$\int_C \pi_{n_1}(x)j_v(x)q(u)\alpha_v(x, x')d(x, u), \tag{3.17}$$

where $C = \{(x, u) : x \in A, x'(x, u) \in B\}$. If the density of (x, u) is $\pi_{n_1}(x)j_v(x)q(u)$ and because we use a transformation of variables (x, u) to x' then the density of the proposal $x'(x, u)$ is equal to

$$\pi_{n_1}(x)j_v(x)q(u)\left|\frac{d(x, u)}{dx'}\right|.$$

This implies that equation (3.17) is equal to

$$\int_{\{x'\in B\}} \pi_{n_1}(x)j_v(x)q(u)\alpha_v(x, x')\left|\frac{d(x, u)}{dx'}\right| dx', \tag{3.18}$$

where now x stands for $x(x')$ and $x' = f(x, u)$. Comparing the above expression with that appearing on the right hand side of equation (3.16), we see that to satisfy detailed

balance condition it is sufficient to satisfy two conditions:

1. There should be a bijection between (x, u) and x'. That is: given the move v', values of X' in B correspond uniquely to values of X in A, and that given move type v and a random value u, the deterministic function f will imply values for $x' \in B$.

2. $$\pi_{n_1}(x) j_v(x) q(u) \alpha_v(x, x') \left| \frac{d(x, u)}{dx'} \right| = \pi_{n_2}(x') j'_v(x') \alpha'_v(x', x). \tag{3.19}$$

   This condition is satisfied through the definition of the acceptance probability $\alpha_v(x, x') = \min\{1, R_v\}$ and $\alpha_{v'}(x', x) = \min\{1, R_v^{-1}\}$ where R_v is as defined in equation (3.14). Using the above definitions, we can deduce

   $$\alpha_v(x, x') R_{v'}(x', x) = \alpha_{v'}(x', x),$$

   which will ensure the condition (1.3).

3.5 Implementing the MCMC algorithm

In implementing the MCMC algorithm, we shall define eight move types : The birth move which will be labelled $v = 1$, the death move labelled $v = 2$, the resize $v = 3$, the shift $v = 4$, changing the angle of rotation move type which will be labelled $v = 5$, updating the intensity level of a cell $v = 6$, the split move $v = 7$ and finally the merge move $v = 8$. We shall assign equal probabilities for choosing the available move types. However, the values of these probabilities will depend on the existing configuration x. For example if there are no cells in the scene, i.e., $n = 0$ and X is an empty set, then the only move type available is the birth move so $j_1(x) = 1$, for $n = 1, j_v(x) = 1/7,\ v = 1, \ldots, 7$ and for $n > 2, j_v(x) = 1/8,\ v = 1, \ldots, 8$.

The object configuration is an unordered set of objects $X = \{X_1, \ldots, X_N\}$. In building up the prior model we started by considering the model for the ordered list of

objects $X = (X_1, \ldots, X_N)$ and modified it accordingly to fit the unordered set. The same concept can be applied in implementing the MCMC moves, however, we found it more convenient to work with the unordered set directly.

3.5.1 Birth and death moves

Suppose our current configuration is $x = \{x_1, \ldots, x_n\} \in \mathcal{U}^n$ and the birth move is selected with probability $j_1(x)$. Then we consider adding a new object x_{n+1} say, which will not overlap with the other objects. We randomly draw a new location $l_{n+1} = (c_{xn+1}, c_{y_{n+1}})$ in the window L, then propose a mark $m_{n+1} = (a_{n+1}, b_{n+1}, \theta_{n+1}, \lambda_{n+1})$ independently. That is we generate a continuous random variable

$$u = (l_{n+1}, a_{n+1}, b_{n+1}, \theta_{n+1}, \lambda_{n+1})$$

from a density $q(u)$. The random variable u has to lie in $\mathcal{U} \subset \mathbb{R}^6$, we have chosen this density $q(u)$ to be the density of a single object defined in equation (3.6), i.e., we draw from the prior.

$$q(u) = f_{X_i}(u) = \pi_l(c_x, c_y)\pi_{a,b}(a_{n+1}, b_{n+1})\pi_\theta(\theta_{n+1})\pi_\lambda(\lambda_{n+1}). \tag{3.20}$$

To deduce x' from x we need to delete one particular cell and this is done through the death move. Given the current state x' the probability of choosing a death move, type labelled 2 is $j_2(x')$. In this move, one cell is selected at random and deleted. As we are dealing with unordered cells, the probability of picking a particular cell among the $n+1$ cells in x' to delete is $1/(n+1)$.

The birth move, if accepted will, yield a configuration $x' = \{x_1, \ldots, x_{n+1}\} \in \mathcal{U}^{n+1}$, a jump to a higher dimension space. Applying the reversible jump algorithm and substituting for the posterior model in equation (3.14), we accept the birth move with

a probability function α_1,

$$\alpha_1(x,x') = \min\left\{1, \frac{\pi(x',(n+1))j_2(x')}{\pi(x,n)q(u)(n+1)j_1(x)} \left|\frac{dx'}{d(u,x)}\right| \frac{\mathcal{L}(y|x',n+1)}{\mathcal{L}(y|x,n)}\right\}. \tag{3.21}$$

If the new object u overlaps with the existing objects in x then $\pi(x',(n+1)) = 0$ and, hence, the acceptance probability $\alpha_1(x,x') = 0$ and we reject the birth move.

The variable u used in moving from x to x' is made up of the parameters of the new cell x_{n+1} so the change of the variable x to x' is trivial, i.e., $x' = x \cup \{u\}$ and the Jacobian is one. The evaluation of the prior distribution for (x,n) in $\mathcal{U}^n$ with respect to $\Gamma_n(x)$ is done by computing $W_{X,N}(x,n)$ defined in equation (3.10). Same steps we follow to compute $\pi(x',n+1)$. In evaluating the ratio of the prior distribution in equation (3.21), most of the terms involved will cancel each other. The remaining term involves the density of x_{n+1}, $f_{X_i}(x_{n+1})$. This density is the same as $q(u)$, hence, will be cancelled. As we have assigned an equal probability for choosing the move types 1 and 2 for n greater than or equal to 1, so $j_2(x')/j_1(x) = 1$. Thus the resulting terms involving the prior, the proposal and the Jacobian are

$$\frac{\beta \mathcal{A}_L}{n+1} I[\text{no overlap for the n+1 objects in } x']. \tag{3.22}$$

For the likelihood ratio $\mathcal{L}(y/x',n+1)/\mathcal{L}(y/x,n)$ we can derive an abbreviated formula as follows

$$\mathcal{L}(y|(x \cup \{x_{n+1}\}),n+1)/\mathcal{L}(y|x,n) = \frac{\prod_{j=1}^{S} \pi(y_j|(x \cup \{x_{n+1}\}),n+1)}{\prod_{j=1}^{S} \pi(y_j|x,n)}$$

$$= \prod_{j \in \Re(x_{n+1})} \frac{\pi(y_j|(x \cup \{x_{n+1}\}),n+1)}{\pi(y_j|x,n)}. \tag{3.23}$$

Equation (3.23) involves all the pixels that are occupied by the object x_{n+1}. The records $Y_j, j = 1,\ldots,S$, are conditionally independent given X, and normally distributed with

mean μ_j and variance σ_j^2 where μ_j and σ_j^2 are functions of $\tau_0, \tau_i, \phi_0, \phi_i$ as defined in section (3.1). Substituting for the likelihood function in equation (2.1), we get

$$\frac{\mathcal{L}(y|x')}{\mathcal{L}(y|x)} = (\frac{\phi_0}{\phi_{x_{n+1}}})^{n_{x_{n+1}}} \exp \sum_{i \in \Re(x_{n+1})} -\frac{1}{2\phi_{x_{n+1}}^2}(y_i - \tau_{x_{n+1}})^2 + \frac{1}{2\phi_0^2}(y_i - \tau_0)^2, \quad (3.24)$$

where $\Re(x_{n+1})$ is the set of indices i such that the signals Y_i contain a contribution from at least one pixel in the object x_{n+1}, $n_{x_{n+1}}$ denotes the number of pixels in $\Re(x_{n+1})$, and $\tau_{x_{n+1}} = \lambda_{n+1} + \tau_0$ is the mean intensity of pixels covered by the cell x_{n+1}.

Multiplying equations (3.22) and (3.24) we get the acceptance probability of a birth that takes the state space from $x = \{x_1, \ldots, x_n\}$ to $x' = \{x_1, \ldots, x_{n+1}\}$ which is equal to

$$\min\{1, R\} \quad (3.25)$$

where

$$R = \frac{\beta \mathcal{A}_L}{n+1}(\frac{\phi_0}{\phi_{x_{n+1}}})^{n_{x_{n+1}}} \exp(\textstyle\sum_{i \in \Re(x_{n+1})} -\frac{1}{2\phi_{x_{n+1}}^2}(y_i - \tau_{x_{n+1}})^2 + \frac{1}{2\phi_0^2}(y_i - \tau_0)^2)$$
$$I[\text{no overlap for the n+1 objects}].$$

If accepted then we move to $x' = x \cup \{x_{n+1}\}$. For the acceptance probability for the corresponding death move some obvious changes to the variables in the ratio terms are made and the ratios are inverted.

3.5.2 The shift move

In the shift move we attempt to update the configuration x by changing the location l of a randomly selected object from $\{x_1, \ldots, x_n\}$, say x_i. Having selected i, updating the location is done by holding the other components fixed and proposing $l_i' = (c_{x_i}', c_{y_i}')$. To propose l_i', we define two independent random variables d_x and d_y, each drawn from a truncated normal distribution with mean zero and variance δ^2. We then set

$l_i' = (c_{x_i}', c_{y_i}')$ to be $(c_{x_i} + d_x, c_{y_i} + d_y)$. To ensure that $l_i' \in L$, we choose d_x and d_y from $[-c_{x_i}, L_x - c_{x_i})$ and $[-c_{y_i}, L_y - c_{y_i})$ respectively, where L_x is the upper $x-$value of the rectangular window L placing the window at the origin $O = (0,0)$ and L_y is the upper $y-$value of L. There is no dimension changing involved in the vector of parameters in the shift move, hence, we can use the standard Metropolis-Hastings algorithm with the following acceptance probability formula

$$\alpha(l_i, l_i') = \min\left\{1, \frac{\pi_{l|\ldots}(l_i')q(l_i', l_i)}{\pi_{l|\ldots}(l_i)q(l_i, l_i')}\right\}. \tag{3.26}$$

Here $\pi_{l|\ldots}(l_i')$ is the conditional distribution of l given all the other parameters in the model (which will be denoted by $|\ldots$), and $q(l_i', l_i)$ is the proposal density of going from state l_i' to state l_i. Since the proposal densities $q(l_i', l_i)$ and $q(l_i, l_i')$ represent truncated normal distributions we need to evaluate their normalising constants to compute the acceptance probability for the move.

Let x_i' be the resulting object after the proposed change, then it follows from equation (3.26) that the acceptance rate for shifting l_i to l_i' can explicitly written as

1. $\alpha(l_i, l_i') = 0$ if the new ellipse x_i' overlaps with the other objects in the scene. Hence the move will be rejected.

2. We accept the move to l_i' with probability

$$\alpha(l_i, l_i') = \min\{1, \frac{z_{l_i'}}{z_{l_i}} \prod_{j \in \Re(x_i) \cup \Re(x_i')} \frac{\sigma_{(j|x_i)}}{\sigma_{(j|x_i')}} \exp(-\frac{1}{2\sigma^2_{(j|x_i')}}(y_j - \mu_{(j|x_i')})^2 + \frac{1}{2\sigma^2_{(j|x_i)}}(y_j - \mu_{(j|x_i)})^2)\}, \tag{3.27}$$

where

$$z_{l_i} = \left[(\Phi(\frac{L_x - c_{x_i}}{\delta}) - \Phi(\frac{-c_{x_i}}{\delta}))(\Phi(\frac{L_y - c_{y_i}}{\delta}) - \Phi(\frac{-c_{y_i}}{\delta}))\right]^{-1},$$

and

$$z_{l'_I} = \left[(\Phi(\frac{L_x - c_{x'_i}}{\delta}) - \Phi(\frac{-c_{x'_i}}{\delta}))(\Phi(\frac{L_y - c_{y'_i}}{\delta}) - \Phi(\frac{-c_{y'_i}}{\delta}))\right]^{-1}.$$

Here L_x is the upper $x-$value of the rectangular window L placing the window at the origin $O = (0,0)$ and L_y is the upper $y-$value of L.

3.5.3 Resizing

If the resize move is selected, then we randomly select an object from $\{x_1, \ldots, x_n\}$, say x_i. Conditioned on x_i we propose a' and b' uniformly over $(a-\zeta, a+\zeta)$ and $(b-\zeta, b+\zeta)$ respectively for some fixed value ζ. The semi-axes are required to lie within a certain range so we condition the proposal set to be in that range as well . The proposal density $q((a_i, b_i), (a'_i, b'_i))$ is therefore

$$\prod_{j=\{a,b\}} \frac{1}{(\min\{\max_{axis}, j_i + \zeta\} - \max\{\min_{axis}, j_i - \zeta\})}.$$

Let $x'_i = (l_i, a'_i, b'_i, \theta_i, \lambda_i)$ be the resulting object after updating the semi-axes a and b, then the acceptance rate for resize move of object x_i is

$$\alpha((a_i, b_i), (a'_i, b'_i)) = min\{1, R\}, \tag{3.28}$$

where

$$R = \frac{q((a'_i, b'_i), (a_i, b_i))}{q((a_i, b_i), (a'_i, b'_i))} \exp(-\frac{1}{2\gamma^2}[(a' - \mu_a)^2 + (b' - \mu_b)^2 - (a - \mu_a)^2$$
$$-(b - \mu_b)^2]) \prod_{j \in \Re(x_i) \cup \Re(x'_i)} \frac{\sigma_{(j|x_i)}}{\sigma_{(j|x'_i)}} \exp(-\frac{1}{2\sigma^2_{(j|x'_i)}}(y_j - \mu_{(j|x'_i)})^2 + \frac{1}{2\sigma^2_{(j|x_i)}}$$
$$(y_j - \mu_{(j|x_i)})^2) I[a, b, a', b' \in M_{axis}] I[\text{no overlap in } x'],$$

where $\Re(x_i) \cup \Re(x'_i)$ is the set of indices j such that the signals Y_j contain a contribution in the object x_i or in the object x'_i, and $\mu_{(j|x_i)}$ and $\sigma^2_{(j|x_i)}$ are the mean and the variance

respectively, of record Y_j conditioned on the existence of object x_i in the configuration. If the semi-axes lie within ζ from the boundaries of M_{axis} then the proposal distribution ratio $q((a_i', b_i'), (a_i, b_i))/q((a_i, b_i), (a_i', b_i'))$ is equal to one. Then the symmetry of the joint proposal distribution implies that R is the ratio of conditional densities of (a', b') and (a, b) under the posterior image distribution.

3.5.4 Orientation

For the re-angling move we draw the proposed candidate θ_i' from the uniform distribution over $(\theta_i - \eta, \theta_i + \eta)$ for some fixed value η. The symmetry of the shape of the ellipse around its axis, implies that a rotation of θ or $\pi + \theta$ produce the same position of the ellipse in L. Hence, to define all the unique positions of a rotated ellipse it is enough to define our state space to be $[0, \pi)$. So if the candidate $\theta_i' < 0$ then we use $\theta_i' = \theta_i' + \pi$, and when $\theta_i' > \pi$ we set $\theta_i' = \theta_i' - \pi$. In the proposal density $U(\theta_i - \eta, \theta_i + \eta)$, assigning small values for η will improve the acceptance rate of the rotation move while large η will propose larger changes in θ_i but the probability of acceptance will be lower. The move is accepted with probability

$$\alpha(\theta_i, \theta_i') = \min\{1, \frac{|\cos(\theta')| + \frac{1}{\pi}}{|\cos(\theta)| + \frac{1}{\pi}} \prod_{j \in \Re(x_i) \cup \Re(x_i')} \frac{\sigma_{(j|x_i)}}{\sigma_{(j|x_i')}} \exp(-\frac{1}{2\sigma^2_{(j|x_i')}}(y_j - \mu_{(j|x_i')})^2 + \frac{1}{2\sigma^2_{(j|x_i)}}(y_j - \mu_{(j|x_i)})^2) I[\text{no overlap between } x_i' \text{ and } \{x|x_I\}\ ]\}. \quad (3.29)$$

3.5.5 Updating the intensity level

To update the intensity level λ of a randomly selected cell $x_i, i \in \{1, \ldots, n\}$, we draw the proposed level λ_i' uniformly over $(\lambda_i - \iota, \lambda_i + \iota)$, where ι has an arbitrary fixed value. As there is no dimension changing involved the standard metropolis-Hastings acceptance probability is used. Analogously to the shift moves, we shall use a truncated uniform distribution to ensure the proposal λ' lies in M_λ. The proposal distributions

for updating the intensity levels are symmetric if both λ_i and λ_i' are within a difference of ι from the boundaries of M_λ, thus the ratio of the proposal densities involved in the acceptance rate for this move, $q(\lambda_i', \lambda_i)|q(\lambda_i, \lambda_i')$ is one. Otherwise this ratio is equal to

$$\frac{(\min\{\max_\lambda, \lambda_i + \iota\} - \max\{\min_\lambda, \lambda_i - \iota\})}{(\min\{\max_\lambda, \lambda_i' + \iota\} - \max\{\min_\lambda, \lambda_i' - \iota\})}.$$

Let x_i' be the modified cell implied by changing λ_i to λ_i' then, the acceptance probability for this move is $\alpha(\lambda_i, \lambda_i') = \min\{1, R\}$ where

$$R = \frac{q(\lambda_i', \lambda_i)}{q(\lambda_i, \lambda_i')}\left(\frac{\phi_i(\lambda)}{\phi_i(\lambda')}\right)^{n_i} \exp \sum_{j \in \Re(x_i)} -\frac{1}{2\phi_i^2(\lambda')}(y_j - \tau_i(\lambda'))^2 + \frac{1}{2\phi_i^2(\lambda)}(y_j - \tau_i(\lambda))^2. \tag{3.30}$$

Unlike the other moves which attempt to update the characteristic of a cell this move will not change $\Re(x_i)$. However, it will change the mean and the variance of the records in the cell x_i.

3.5.6 Split and merge moves

As well as the previous moves, we define two extra transitions, split and merge, which are useful when there is uncertainty about whether the existing object actually covers two close objects or vice versa. We define two objects to be ϵ-close if the Euclidian distance between their centres is less than some specified value ϵ_d, the difference of their intensity does not exceed some fixed value ϵ_λ and the difference between their angle of rotation is less than ϵ_θ, i.e.,

$$\sqrt{(c_{xA} - c_{xB})^2 + (c_{yA} - c_{yB})^2} < \epsilon_d,$$

and

$$|\theta_A - \theta_B| < \epsilon_\theta \text{ or } |\theta_A - \theta_B| > \pi - \epsilon_\theta,$$

and

$$|\lambda_A - \lambda_B| < \epsilon_\lambda.$$

If the data are reasonably good (as we believe in this experiment) and the objects are close, then using the death and birth moves alone, intermediate configurations between the one and the two object states will have prohibitively low probability. For example, if the existing object(s) cover(s) mostly the object areas and have a reasonable intensity level(s), then the chance for these object(s) to change their positions or sizes or any of their mark variables will be low, and moreover they are unlikely to be deleted. Hence, the probability of changing the configuration from the one cell to the two cells or vice versa through a sequence of move labels $v = 1, \ldots, 6$ will be very low.

We shall briefly describe the mechanism of the merge and split moves. For the merge move we attempt to combine two randomly selected objects say, x_A, x_B if they are ϵ-close. We propose replacing them by a new object, say x_C (see figure $(3-7)$). The combining operation must be done such that the reverse splitting move can be constructed to satisfy the detailed balance condition. In order for the proposed move to be accepted, the area covered by the object C, $\mathcal{A}_{x_C}$ should be almost equal to $\mathcal{A}_{x_A} \cup \mathcal{A}_{x_B}$. To do this we allocate the centre of the new merged object, $l_C = (c_{xC}, c_{yC})$ on the centre of l_A and l_B, i.e.,

$$c_{xC} = \frac{c_{xA}+c_{xB}}{2} \qquad c_{yC} = \frac{c_{yA}+c_{yB}}{2} \tag{3.31}$$

There are several qualitatively different positions the two cells A and B can take when they are proposed to merge. Figure $(3-7)$ displays some of these positions. It is clear that the semi-axes and the angle of rotation of the new single object must be chosen carefully for a proposal to merge to be acceptable. We will consider proposing a convenient semi-axes for the new cell C conditioned on the location of the centres l_A and l_B. Let Δ be the *acute* angle of a right triangle obtained from the intersection of

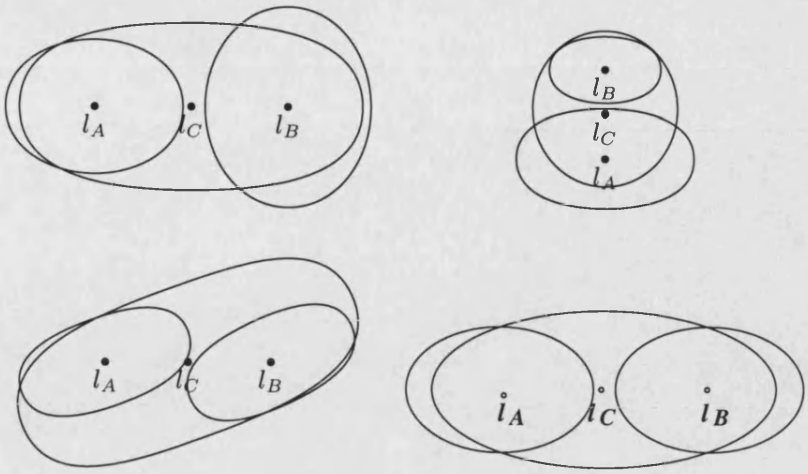

Figure 3-7: Different positions for two cells to be merged

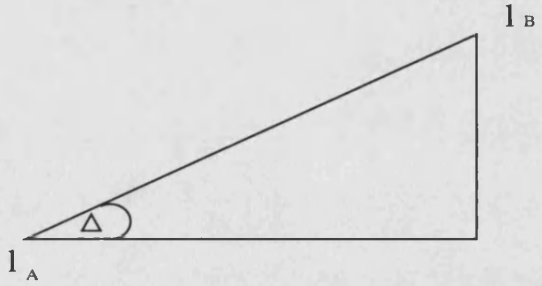

Figure 3-8: The acute angle Δ of a right triangle.

the hypotenuse side ($\overline{l_A l_B}$) and the adjacent side to the right angle (see figure $(3-8)$).

$$\Delta = \arctan \frac{c_{yB} - c_{yA}}{c_{xB} - c_{xA}}.$$

To sketch values of semi-axis suitable for the merged cell C to cover an appropriate area for different values of Δ, we consider the different cases shown in figure $(3-7)$. The first situation is shown in the upper left corner and the lower right corner of figure $(3-7)$ where $\Delta = 0$ meaning that the two cells' centres coincide in their $y-$coordinate value. That is the merged cell can cover up to the sum of the two semi-axes in an

approximate direction of $x-$axis bearing in mind the rotation angles of most of the cells are close to either zero or π. Hence, a good value to propose for a_C is the sum of a_A and a_B. Whereas for the other semi-axes, b of the merged cell C, b_C, will be limited by the maximum of b of the two existing cells A and B. Therefore we can approximate b_C to be the average value of b_A and b_B. For the other extreme case where $\Delta = \pi/2$, displayed in the upper right corner of figure $(3-7)$, a good proposal value for a_C is $(a_A + a_B)/2$ while a reasonable proposal for b_C is $b_A + b_B$. In the intermediate case $\Delta = \pi/4$, sensible proposals for a_C and b_C are $(a_A+a_B)/2$ and $(b_A+b_B)/2$ respectively. Consequently we conclude that the value of the semi-axes of the proposed cell should be associated with the angle Δ. The plausible proposals identified in the three cases just considered can be achieved by setting

$$a_C = \frac{1}{2}(1 + |\cos(\Delta)|)(a_A + a_B),$$

$$b_C = \frac{1}{2}(1 + |\sin(\Delta)|)(b_A + b_B),$$

providing that $a_C, b_C \in (\min_{axis}, \max_{axis})$ as will be verified later. The above approach to determining the semi-axes holds for small angles of rotation θ. For small values of θ or values close to π, where the major semi-axes lie approximately in the direction of the $x-$ axis and the minor semi-axis lie in the direction of the $y-$axis. A general formula can be derived using the same concept for other situations where ellipses tends to lie in a different preferred direction. The previous conditions of having the resulting semi-axes in the required space will prohibit merging big cells if their corresponding semi-axes were close to the upper limit of the semi-axes space, i.e., the merge shall be proposed if

$$a_A + a_B \in (\frac{2\min_{axis}}{|\cos(\Delta)|+1}, \frac{2\max_{axis}}{|\cos(\Delta)|+1}),$$

$$\text{and} \qquad b_A + b_B \in (\frac{2\min_{axis}}{|\sin(\Delta)|+1}, \frac{2\max_{axis}}{|\sin(\Delta)|+1}). \tag{3.32}$$

To find the angle of rotation θ_C and the intensity λ_C, the merge move is more likely to be deterministic, as we define θ_C to be the average value of θ_A and θ_B, and similarly, the intensity level of the new object is $\lambda_C = (\lambda_A + \lambda_B)/2$.

For the reverse splitting move, we start with a cell $x_C = (l_C, a_C, b_C, \theta_C, \lambda_C)$ and propose to split it into two cells x_A and x_B. In order to have a bijection between the values of x_C and the values of x_A, x_B to attain detailed balance, the proposal mechanism of choosing x_A and x_B has to be set up carefully. First, we need to define two centres l_A, l_B satisfying the equation $l_C = (l_A + l_B)/2$. This is implemented by drawing one centre say l_A uniformly in a disc centred on l_C with diameter ϵ_d, and letting this determine the centre of the other object B so we have $l_C = (l_A + l_B)/2$. We draw Δ, the acute angle of l_A and l_B uniformly over $(0, 2\pi)$, and also r uniformly over $(0, \frac{\epsilon_d}{2})$, and set

$$\begin{aligned} c_{xA} &= c_{xC} + r\cos(\Delta) & \qquad c_{yA} &= c_{yC} + r\sin(\Delta) \\ c_{xB} &= c_{xC} + r\cos(\Delta + \pi) & \qquad c_{yB} &= c_{yC} + r\sin(\Delta + \pi) \end{aligned} \tag{3.33}$$

where (c_{xi}, c_{y_i}) is the $x-$coordinate and the $y-$coordinate of l_i. First of all we require r to be $\leq \epsilon_d/2$, but we also want l_i to lie in L so we shall put more constraints on the range of r. The choice of the radius r will be conditioned on Δ. The lower limit of r is bounded by $\min_{axis}$, because of the non-overlapping condition; two cells in the window L should be at least $2\min_{axis}$ distance apart not to overlap. The upper bound, $\max_r$, of the radius r can be computed conditioned on Δ as follows:

1. If $0 < \Delta < \frac{\pi}{2}$, then $\max_r = \min\{\frac{L_x - c_{xC}}{\cos(\Delta)}, \frac{c_{xC}}{\cos\Delta}, \frac{L_y - c_{yC}}{\sin(\Delta)}, \frac{c_{yC}}{\sin(\Delta)}, \frac{\epsilon_d}{2}\}$.

2. If $\frac{\pi}{2} < \Delta < \pi$, then the radius r will be less than

$$\max_r = \min\{\frac{-L_x + c_{xC}}{\cos(\Delta)}, \frac{-c_{xC}}{\cos\Delta}, \frac{L_y - c_{yC}}{\sin(\Delta)}, \frac{c_{yC}}{\sin(\Delta)}, \frac{\epsilon_d}{2}\}.$$

3. If $\pi < \Delta < \frac{3\pi}{2}$, then $\max_r = \min\{\frac{-L_x + c_{xC}}{\cos(\Delta)}, \frac{-c_{xC}}{\cos\Delta}, \frac{-L_y + c_{yC}}{\sin(\Delta)}, \frac{-c_{yC}}{\sin(\Delta)}, \frac{\epsilon_d}{2}\}$.

4. If $\frac{3\pi}{2} < \Delta < 2\pi$, then the radius r will be less than

$$\max_r = \min\{\frac{L_x - c_{xC}}{\cos(\Delta)}, \frac{c_{xC}}{\cos\Delta}, \frac{-L_y + c_{yC}}{\sin(\Delta)}, \frac{-c_{yC}}{\sin(\Delta)}, \frac{\epsilon_d}{2}\}.$$

Determining the semi-axes for the new split cells should be done such that there is a bijection between the set of values of (a_C, b_C) and the set of values of $(a_A, b_A), (a_B, b_B)$. Thus we generate two random variables u_a, u_b from a truncated normal distribution with mean zero and variance ϖ^2, and set

$$\begin{aligned} a_A &= \frac{a_C}{|\cos(\Delta)| + 1} + u_a, \qquad a_B = \frac{a_C}{|\cos(\Delta)| + 1} - u_a \\ b_A &= \frac{b_C}{|\sin(\Delta)| + 1} + u_b, \qquad b_B = \frac{b_C}{|\sin(\Delta)| + 1} - u_b. \end{aligned} \tag{3.34}$$

The condition that the axes have to lie within M_{axis} implies that

$$u_a \leq \max_{u_a} = \min\{\max_{axis} -\frac{a_C}{|\cos(\Delta)| + 1}, -\min_{axis} + \frac{a_C}{|\cos(\Delta)| + 1}\},$$

and

$$u_a \geq \min_{u_a} = \max\{\min_{axis} -\frac{a_C}{|\cos(\Delta)| + 1}, -\max_{axis} + \frac{a_C}{|\cos(\Delta)| + 1}\},$$

where these boundaries are imposed at the sampling stage to ensure that the proposed semi-axes lie in their defined spaces. Similarly the lower bound $\min_{u_b}$ and the upper bound $\max_{u_b}$ of u_b are computed and are equal to

$$\max\{\min_{axis} -\frac{b_C}{|\sin(\Delta)| + 1}, -\max_{axis} + \frac{b_C}{|\sin(\Delta)| + 1}\} \text{ and } -\min_{u_b}$$

respectively. The bijection of the semi-axes values in the merge and split moves implies that the split move is accepted if the conditions of (3.32) is satisfied. The choice of the parameters of distribution of u_a and u_b controls the difference between the relative semi-axes of the two split cells. By choosing their mean to be zero and using small

variances we encourage the two cells to have approximately equal axes, i.e., split C into two equally sized cells A and B.

For the angles of rotation and the intensity levels we draw two random variables u_θ uniformly from $(-\frac{\epsilon_\theta}{2}, \frac{\epsilon_\theta}{2})$ and u_λ from a uniform distribution such that $\min_{u_\lambda} \leq u_\lambda \leq \max_{u_\lambda}$. We set

$$\begin{aligned} \theta_A &= \theta_C - u_\theta \qquad \theta_B = \theta_C + u_\theta, \\ \lambda_A &= \lambda_C - u_\lambda \qquad \lambda_B = \lambda_C + u_\lambda. \end{aligned} \tag{3.35}$$

To ensure that λ_A, λ_B are in the range $(\min_\lambda, \max_\lambda)$, we condition u_λ to be bounded by $\min_{u_\lambda} = \max\{-\frac{\epsilon_\lambda}{2}, \lambda_C - \max_\lambda, \min_\lambda - \lambda_C\}$ and $\max_{u_\lambda} = \min\{\frac{\epsilon_\lambda}{2}, \lambda_C - \min_\lambda, \max_\lambda - \lambda_C\}$.

The merge (or split) move if accepted will reduce (or increase) the dimension of the current configuration by one. So Green's algorithm defined in section (3.4) is applied here. Suppose that the current state is $x = \{x_1, \ldots, x_{n-1}, x_n, x_{n+1}\}$ and we propose to combine two randomly selected cells say x_n, x_{n+1} to x_n^* which if accepted will take the current state to $x' = \{x_1, \ldots, x_{n-1}, x_n^*\}$. We accept the merge move with a probability function α_7, which is maintained by applying the reversible jump algorithm and substituting for the posterior model,

$$\alpha_7(x, x') = \min\left\{1, \frac{\pi(x', n) j_8(x') q(u)(n+1)}{\pi(x, n+1) j_7(x) 2} \left|\frac{d(u, x)}{dx'}\right| \frac{\mathcal{L}(y|x', n)}{\mathcal{L}(y|x, n+1)}\right\}. \tag{3.36}$$

The basic terms involved in calculation of the acceptance probability are as follows

1. The first term, from the prior contribution to $\mathcal{W}_{X,N}(x', n)/\mathcal{W}_{X,N}(x, n+1)$, is

$$\frac{1}{\beta} \frac{\pi_{a,b}(a_n^*, b_n^*)}{\pi_{a,b}(a_n, b_n)\pi_{a,b}(a_{n+1}, b_{n+1})} \frac{\pi_\theta(\theta_n^*)}{\pi_\theta(\theta_n)\pi_\theta(\theta_{n+1})} \frac{\pi_\lambda(\lambda_n^*)}{\pi_\lambda(\lambda_n)\pi_\lambda(\lambda_{n+1})} I[\text{no overlap between the existing objects}], \tag{3.37}$$

which is equal to

$$\frac{12z_1^{-1}\pi\gamma^2(|\cos(\theta_n^*)|+\frac{1}{\pi})(\max_\lambda-\min_\lambda)}{\beta(|\cos(\theta_n)|+\frac{1}{\pi})(|\cos(\theta_{n+1})|+\frac{1}{\pi})}\exp-\frac{1}{2\gamma^2}((a_n^*-\mu_a)^2+$$
$$(b_n^*-\mu_b)^2-(a_n-\mu_a)^2-(b_n-\mu_b)^2)-(a_{n+1}-\mu_a)^2-(b_{n+1}-$$
$$\mu_b)^2)I[\text{no overlap}]I[\min_{axis}\leq a_n^*,a_n,a_{n+1},b_n^*,b_n,b_{n+1}\leq\max_{axis}], \qquad (3.38)$$

where z_1 is the normalising constant for the density of the semi-axes and is equal to

$$[\int_{\min_{axis}}^{\max_{axis}}\int_{\min_{axis}}^{\max_{axis}}\frac{1}{2\pi\gamma^2}\exp(-\frac{1}{2\gamma^2}((x-\mu_a)^2+(y-\mu_b)^2))dxdy]^{-1}.$$

2. The second term, $j_8(x')q(u)(n+1)/2j_7(x)$, is

$$\frac{n+1}{2}\frac{j_{split}(x')}{j_{merge}(x)}q_{u_\lambda}(u_\lambda)q_{u_\theta}(u_\theta)q_\Delta(\Delta)q_{r|\Delta}(r)q_{u_a}(u_a)q_{u_b}(u_b). \qquad (3.39)$$

where the term $(n+1)/2$ is from the ratio of probability of picking up the correct cell to split in the state x', which is equal to $= 1/n$, over the probability of chosen the required cells from x to be merged $= \begin{pmatrix} n+1 \\ 2 \end{pmatrix}^{-1}$. Substituting the specified proposal distributions, the proposal ratio becomes

$$\frac{(n+1)z_2}{8\pi^2\varpi^2\epsilon_\theta(\max_{u-\lambda}-\min_{u-\lambda})\max_r}\exp(-\frac{u_a^2+u_b^2}{2\varpi^2})$$
$$I[\min_{u_a}\leq u_a\leq\max_{u_a},\min_{u_b}\leq u_b\leq\max_{u_b}],$$

where

$$z_2=[\int_{\min_{u_b}}^{\max_{u_b}}\int_{\min_{u_a}}^{\max_{u_a}}\frac{1}{2\pi\varpi^2}\exp-\frac{x^2+y^2}{2\varpi^2}dxdy]^{-1}.$$

3. The Jacobian of the transformation from

$$(l_n^*,r,\Delta,a_n^*,u_a,b_n^*,u_b,\theta_n^*,u_\theta,\lambda_n^*,u_\lambda)$$

to

$$(l_n, l_{n+1}, a_n, b_n, a_{n+1}, b_{n+1}, \theta_n, \theta_{n+1}, \lambda_n, \lambda_{n+1})$$

using equations (3.33), (3.34) and (3.35) is

$$\frac{(|\cos(\Delta)|+1)(|\sin(\Delta)|+1)}{64r}.$$

4. The likelihood ratio contribution $\mathcal{L}(y|x',n)/\mathcal{L}(y|x,n+1)$ involves all the pixels that are currently occupied by x_n, x_{n+1} in addition to the pixels that will be covered by cell x_n^*, that is it will involve the set of indices in $\Re(x_n) \cup \Re(x_{n+1}) \cup \Re(x_n^*)$. The likelihood ratio can be basically written as a product of three different ratios :

   - The likelihood of the pixels that are covered by x_n, x_{n+1} and will be covered by x_n^*,

$$\prod_{k=n}^{n+1} \prod_{j \in \Re(x_k) \cap \Re(x_n^*)} \frac{\phi_{x_k}}{\phi_{x_n^*}} \exp\left(-\frac{1}{2\phi_{x_n^*}^2}(y_j - \tau_{x_n^*})^2 + \frac{1}{2\phi_{x_k}^2}(y_j - \tau_{x_k})^2\right). \quad (3.40)$$

   - The likelihood of the pixels which were covered by x_n, x_{n+1} and will not be covered by x_n^*

$$\prod_{k=n}^{n+1} \prod_{j \in \Re(x_k) | \Re(x_k) \cap \Re(x_n^*)} \frac{\phi_{x_k}}{\phi_0} \exp\left(-\frac{1}{2\phi_0^2}(y_j - \tau_0)^2 + \frac{1}{2\phi_{x_k}^2}(y_j - \tau_{x_k})^2\right) \quad (3.41)$$

   - the likelihood of the pixels that are considered as a background in configuration x and falls inside object x_n^* in the proposed state x'

$$\prod_{k=n}^{n+1} \prod_{j \in \Re(x_n^*) | \Re(x_k) \cap \Re(x_n^*)} \frac{\phi_0}{\phi_n^*} \exp\left(-\frac{1}{2\phi_{x_n^*}^2}(y_j - \mu_{x_n^*})^2 + \frac{1}{2\phi_0^2}(y_j - \tau_0)^2\right) \quad (3.42)$$

Combining the four main terms we get R. The acceptance probability for the

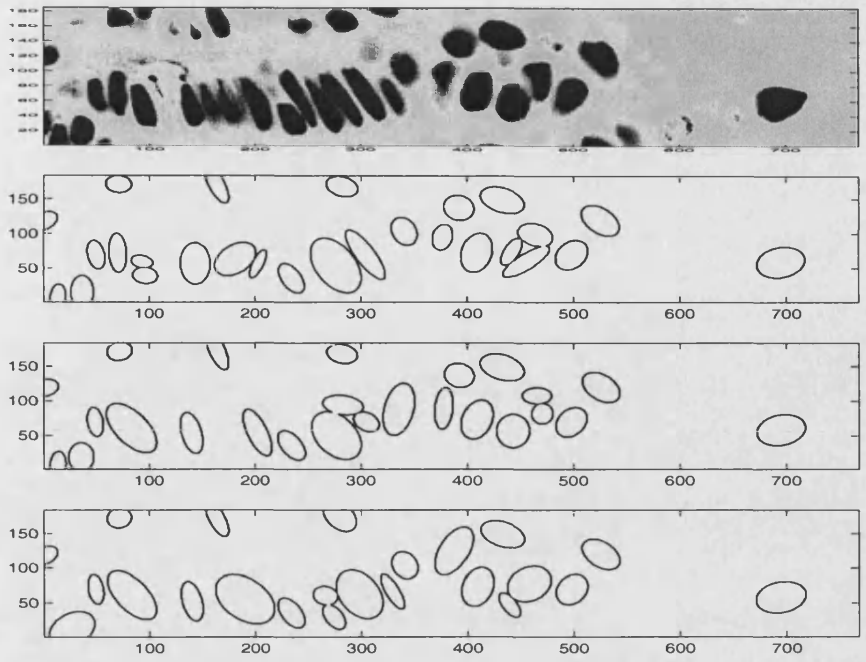

Figure 3-9: The input image (top) and the 20000th sample configurations using different runs.

merge move is $\min\{1, R\}$ and for the split move that will take the state from x' to x is $\min\{1, R^{-1}\}$.

3.6 Sampler performance

Based on a visual inspection of the size, intensity and shape of the cells in the image, we chose the model hyper-parameters as follows: $M_{axis} = [3, 60], M_{\lambda} = [40, 217]$. We expect to have approximately 30 objects in the scene, hence, we set $\beta = 30/\mathcal{A}_L$. For the parameters involved in the prior distribution of the semi-axes (a, b) we choose $\mu_a = 25.0, \mu_b = 12.5, \gamma = 5.5$. We set the scale constant involved in the definition of the variance of the likelihood function to be $c^* = 1$. Focusing on the assumptions of the proposal kernels of the shift, resize, orientation and updating intensity moves we chose $\delta = 2.0, \zeta = 2.0, \eta = \pi/10.0, \iota = 10$. The constants involved in the split and merge moves are assigned the following values: $\epsilon_d = 2\mu_a$, $\epsilon_\theta = \pi/40$, $\epsilon_\lambda = (\max_\lambda - \min_\lambda)/5, \varpi = 3.0$.

Starting with an empty configuration, we ran the sampler for 20000 iterations. Figure $(3-9)$ shows the resulting sample images which allocated some cells in the correct places (comparing it with the input image) but still need more iterations to adjust the position of the other cells. The ellipses in the 2D output samples in this thesis are plotted using a binary grid technique to be consistent with the input data. We set background pixels to zero and pixels in the cells to one and use a contour method to outline the ellipses. It should be noted that by using this technique some adjacent cells may look as though they are one connected set of pixels , however, they do not overlap.

Further investigations have been implemented to study the convergence of the model using these three samples obtained using different runs. To look at convergence we plotted the trace of the likelihood. In other situations, it would be reasonable to look at the posterior density throughout the run time, however this is not feasible in this case. The posterior is evaluated as a collection of different densities, each is defined individually and with respect to different Lebesgue measure on different dimension subspaces. Therefore, the trace of the observed posterior image density implemented in the way described above will not be appropriate to use as a diagnostic for convergence.

Figure (3-10) assesses the likelihood in a logarithmic scale for the three different runs mentioned earlier. There is a reasonable convergence within any one run of the posterior model after a burn-in period. However, different representations in term of cells and their locations were obtained for these modal regions as seen in figure $(3-9)$. The trend of the likelihood for each one run indicates the smooth convergence of the chain. It does not suggest any difficulty in the convergence, in contrary to the previous conclusions obtained from visual inspections of the output configurations. Hence we can argue here that the trace of the likelihood as a main diagnosis for the convergence may be misleading. Linking the visual inspection of the output configurations with the trace of the likelihood we can figure out that within each individual run we have reached a local mode. Moreover, to reach the global mode with the current rate of mixing of

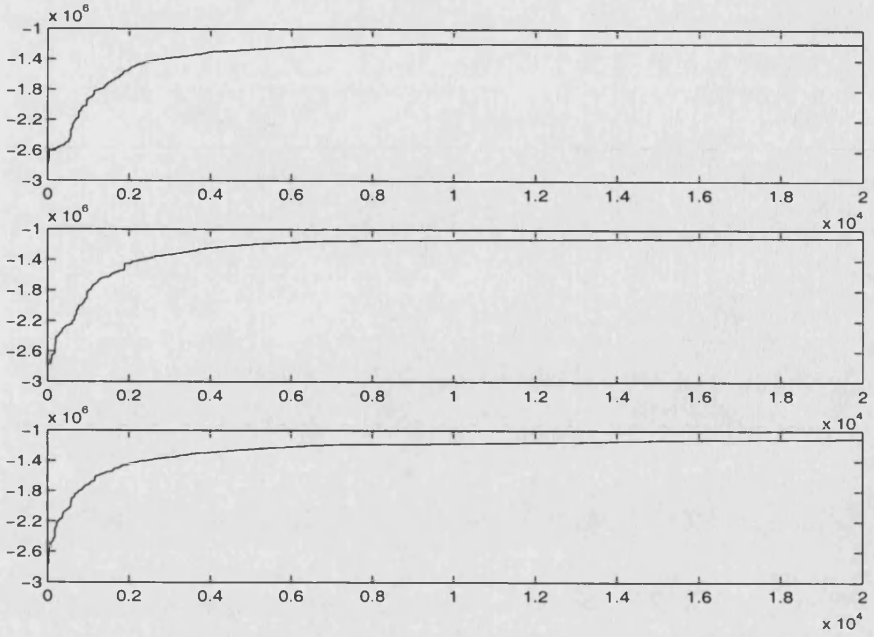

Figure 3-10: The trace of the likelihood for different runs. The vertical axis represents the observed value in logarithmic scale and the horizontal axis is the number of iterations.

the sampler a longer run time is needed. Further, we can say that the posterior image distribution which covers an enormous state space, has several local modes and a highly global mode.

To investigate the mixing of the sampler, we trace traces the cumulative accepted moves with time recorded every 250 sweeps for one of the runs. The results are displayed in figure $(3-11)$. Birth moves are accepted frequently at first, then much more slowly after that. Shift, resize, rotate and intensity acceptances remain frequent throughout the run time as the details for each cell oscillate within the posterior distribution. They are more common just after a cell has been created. The rate of acceptance of these moves in late iterations is lower than it is just after the creation of a new cell. This is because of the random proposals of the location and the marks of a new cell which will imply an acceptance of a cell not necessary fitting the true cell exactly but fitting the data reasonably well. The adjustment proposals toward a better fit are accepted in the following iterations until the cell reaches a good fit to the data. Once it is

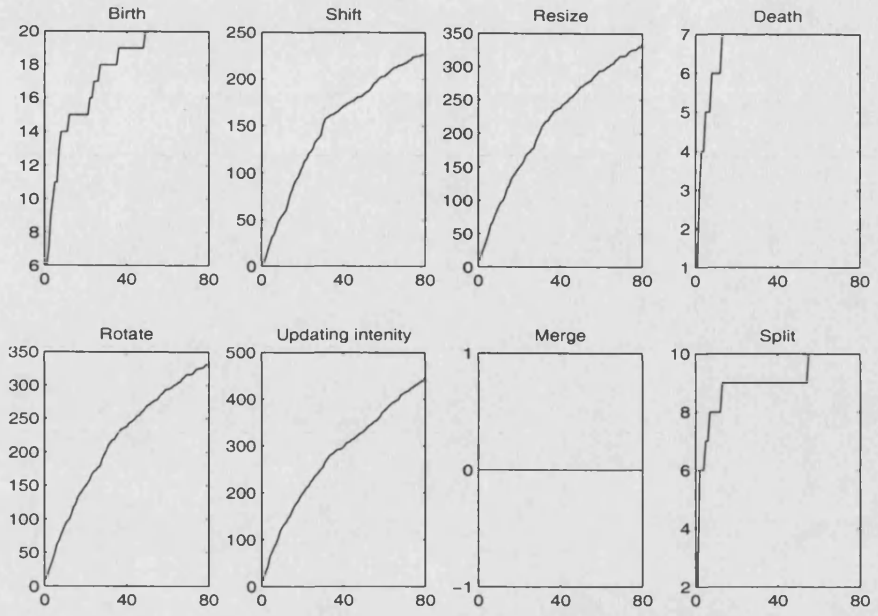

Figure 3-11: A trace of the total number of accepted moves recorded every 250 iterations, where the horizontal axis represents the number of iterations over 250 and the vertical axis is the number of accepted moves of a given type.

there, there are not major improvements to make. For the death move we observed that as we started with an empty configuration, the cells which appeared are added when there is a high acceptance rate R in the formula $\min\{1, R\}$ defined in equation (3.25). So according to the reversible jump concept, the acceptance probability for the reverse death move is $\min\{1, R^{-1}\}$, i.e., the acceptance rate R^{-1} is low. The accepted death moves correspond to cells which resulted from split moves. No merge move was accepted in this run, this may be because of the conditions required for two cells to be merged which are ϵ_d-close, $\epsilon_\theta-$close and $\epsilon_\lambda-$close as defined in section (3.5.6). The values of these ϵs exhibit tension between the merge and the split moves. High values of the ϵs will increase the chance of merging cells and reduce the probability of splitting a cell and vice versa. For our case we need to enhance the chance of splitting the cells rather than merging as figure $(3-9)$ indicates. We may also increase the chance of the merge move by defining a purely representational move. An ellipse with a first semi-axis a, second semi-axis b and angle θ is identical to the ellipse of the same centre

with first semi-axis b, second semi-axis a and angle $\theta + \pi/2$. In our representational move we propose to replace the current (a, b, θ) by $(b, a, \theta + \pi/2)$. This proposal will be accepted with a probability rate

$$\min\left\{1, \frac{\pi_\theta(\theta + \frac{\pi}{2})\pi_{a,b}(b,a)}{\pi_\theta(\theta)\pi_{a,b}(a,b)}\right\}.$$

This move will help in increasing the probability of two close cells to satisfy the required conditions of the merge move: the $\epsilon_\theta-$close condition and the semi-axes conditions stated in equation (3.32).

The likelihood ratio dominates all the other ratios involved in the calculation of the acceptance probability for the move types defined in the algorithm. This is due to the strong data records, their high variation and the size of the cells. More analysis will be addressed to the parameter of the likelihood function c^*, as will be explained in section (3.8).

3.7 Improving the birth move of the algorithm.

The prior distribution for the location of the centres of the cells is assumed to be uniform over the window L, and the proposal mechanism draws a point l uniformly over this window. As background areas in L are bigger than areas in L covered by cells this implies that a high proportion of proposals for births of new cells are made in background areas where the proposed birth is rejected. To increase the acceptance rate for the birth process, another technique for proposing a location of a centre of a new cell in the birth move was examined. We alter the proposal mechanism to be associated with the records $Y = (Y_i; i \in \{1, \ldots, S\})$ such that pixels with high record values are more likely to be the proposed centre l of an object. Instead of drawing the proposed l uniformly over L, we choose the centre of a proposed cell as follows:

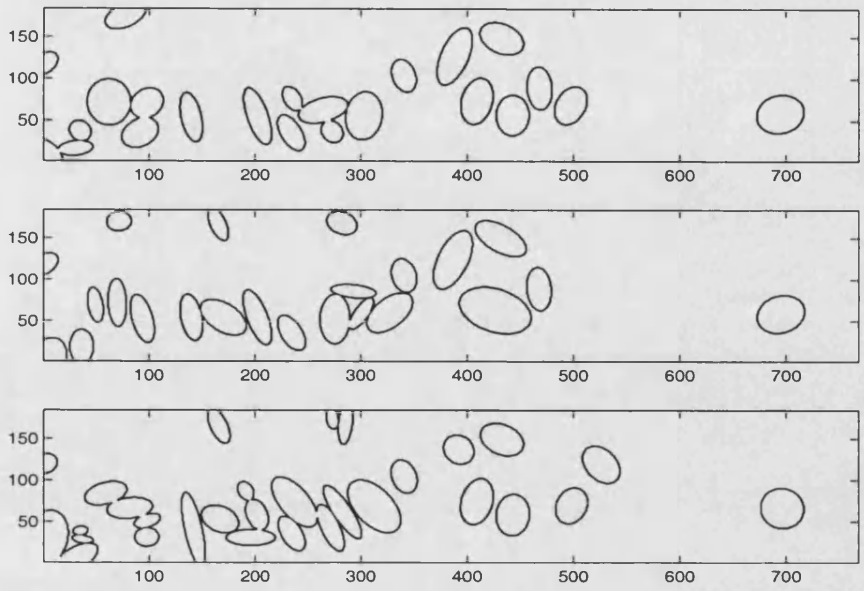

Figure 3-12: The resulting sample images after improving the algorithm. Each corresponds to the 20000 th iteration using different runs.

- choose a pixel i according to the probability distribution, $p_i, i = 1, \ldots, S$ where

$$p_i \propto \begin{cases} 1 & Y_i < 30 \\ \frac{1}{3}Y_i - 9 & 30 \leq Y_i < 90 \\ 21 & 90 \leq Y_i, \end{cases}$$

- then, choose a point uniformly from the region occupied by this pixel.

A note should be made of the pixel i and its associated probability p_i for the computation of the acceptance probability for the birth move as well as the reverse death move. Suppose the current configuration is $x = \{x_1, \ldots, x_n\}$ and we are considering adding a new object x_{n+1}. Let i denote the pixel covering the centre point l. The only change in the density of the proposed cell x_{n+1} is concerned with this location point. The new location l_{n+1} is proposed with probability p_i instead of $\pi_l = 1/\mathcal{A}_L$. Hence, the proposal density of the new object $q(u) = q(x_{n+1})$ defined in equation (3.20) will be $p_i f_{X_i}(u)/\pi_l$. Thus the probability of acceptance of a birth move defined in equation

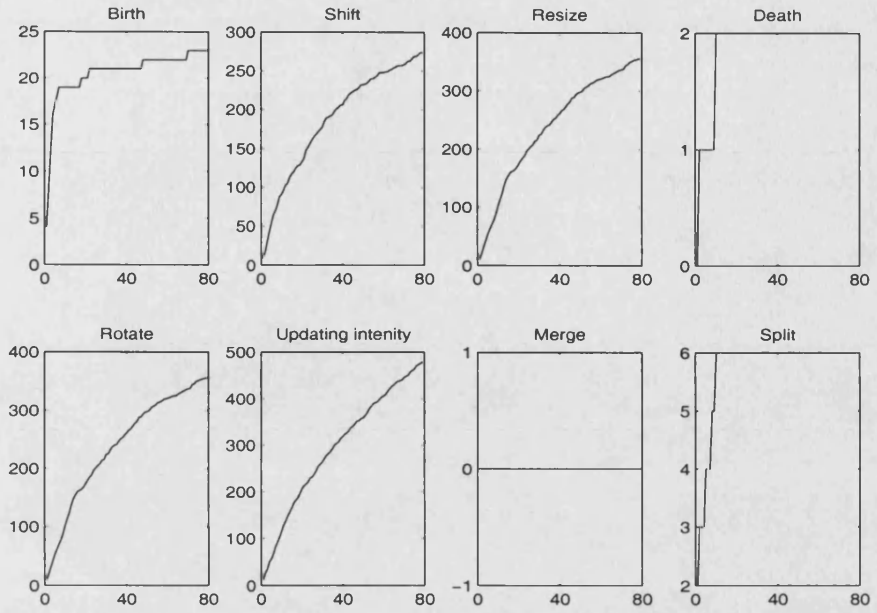

Figure 3-13: The trace of the number of accepted moves recorded every 250 iterations after improving the algorithm.

(3.25) will change to

$$\alpha_1(x, x') = \min\{1, R\}, \tag{3.43}$$

where,

$$R = \frac{\beta}{(n+1)p_i}\left(\frac{\phi_0}{\phi_{x_{n+1}}}\right)^{n_{x_{n+1}}} \exp\Big(\sum_{j \in \Re(x_{n+1})} -\frac{1}{2\phi^2_{x_{n+1}}}(y_j - \tau_{x_{n+1}})^2 + \frac{1}{2\phi_0^2}(y_j - \tau_0)^2\Big) I[\text{no overlap for the n+1 objects}]$$

and for the reverse move will be $\min\{1, R^{-1}\}$ after changing some obvious terms in R.

In our modified algorithm all the other moves for updating the current configuration are implemented as before. The value of R in the acceptance probability of equation (3.43) using the new proposal kernel is now lower for high record pixels and higher for low record pixels than in equation (3.25). However, this change does not affect the probability of acceptance of a birth move if the proposed move fits the data well

enough that the high likelihood ratio still leads to a value $R > 1$. Figure $(3-13)$ traces the total number accepted of each move type recorded every 250 iterations for the modified algorithm for a single run. A comparison between this result and the previous one recorded in figure $(3-11)$ shows that the rate of the accepted birth move at early iterations is higher using the modified algorithm. In later iterations, because there are almost no more "undiscovered" cells, this rate is declining faster than it is for the standard technique.

Comparing the sample output obtained using the standard proposal technique (see figure $(3-9)$) with that obtained using the above technique (see figure $(3-12)$) we notice that the former samples produced larger cells whereas the latter samples contains several small close cells instead of big cells. Using the updated algorithm, the birth of cells is directed mostly to the cells' area. The chance for these cells to update themselves before the birth of a new cell close by is reduced. Once these close cells are allocated then the time for these cells to adapt to the true scene will be long considering the non-overlapping condition and the size of the state space of all the configurations. So the total number of the cells using the improved algorithm is slightly more than it is using the previous algorithm.

The new technique for proposing the location of a new cell has better overall acceptance rate than the old one but it also has an unanticipated adverse effect on convergence. To avoid this problem and gain full advantage of this improved technique in reducing the convergence time, we rearrange the selection of the move scheme for updating the configurations. We assign lower probability for picking the birth move $(j_1(x) = 1/22)$ than for the other move types $j_v(x) = 3/22,\ v = 2,3,4,5,6,7,8$. In this case we have rescheduled the total run time so it will produce a better result as seen in figure $(3-14)$.

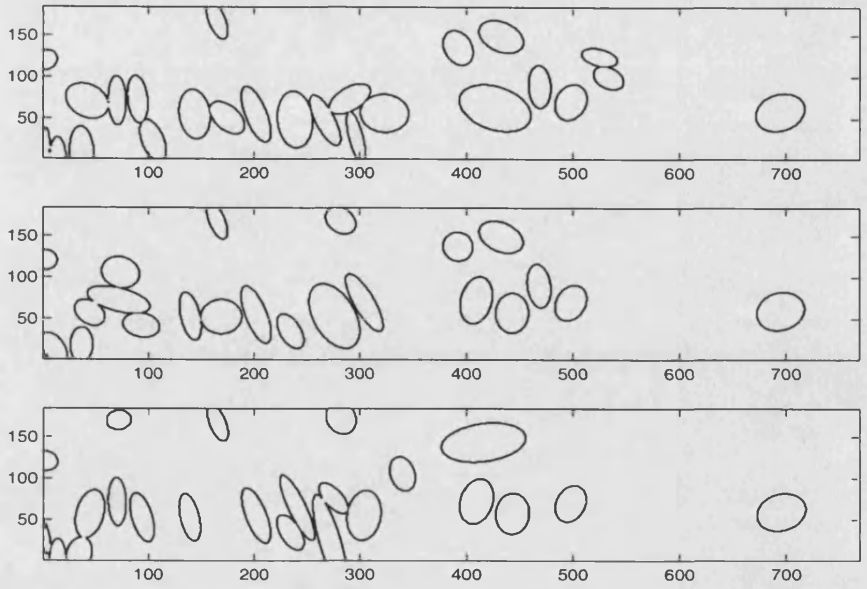

Figure 3-14: The resulting sample images after improving the algorithm and changing the probabilities of picking the move types.

3.8 Parameters of the likelihood function

In order to address some aspects of sensitivity analysis, we examined several choices for the hyper-parameters involved in the target model. Experimental runs indicate that the parameters of the likelihood function are the most influential among the terms involved in the target model. Our model is very sensitive to the values of τ_0, ϕ_0 and c^*. By fixing the values of τ_0, ϕ_0 to the ones obtained from data analysis, we have employed several values for the scale parameter c^*. The values of $c^* > 3.0$ are a potential source of instability; using high values of c^*, such as 3 and 4.5, the sampler classified wrongly some background with relatively high records as cells or parts of cells. With the smaller values $c^* = 0.5$ and $c^* = 1.0$, the resulting cells seem to match the input image better (see figures $(3-14)$ and $(3-15)$ for $c^* = 1.0$ and 0.50 respectively). Using small values for c^* produces smaller cells variances than when using large values and hence reduces the chance of classifying high background pixels as cells or part of cells.

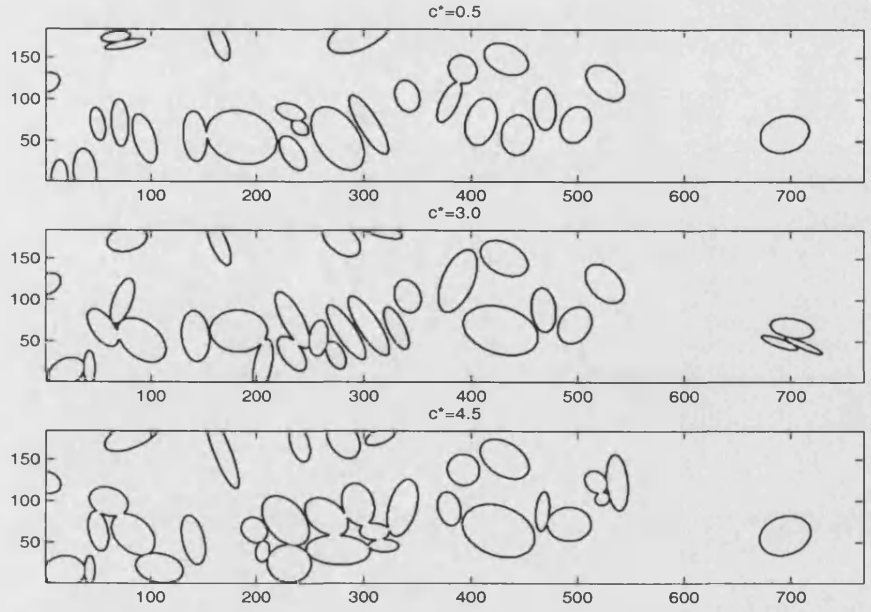

Figure 3-15: The resulting sample images using $c^* = 0.5$ (top), $c^* = 3.0$ (middle) and $c^* = 4.5$ (bottom).

3.8.1 Modifying parameters of the likelihood model

In the previous examples we observed that the samples have managed to locate most of the cells in the correct places, however, some cells need adjustments toward a better fit of the data. The likelihood function is sensitive to the value of c^* which affects the variance of the Gaussian distribution of the records Y. A value of 1 was set for c^* to reflect a good interpretation of the relation between the sample means and sample variances for small intensity cells as described in section (3.1). Further analysis for the cells was carried out with a view to modifying our model for the intensity level of a cell. We observed the record values for the pixels falling inside some selected cells. Figure $(3 - 16)$ is a 3D plot for the records of three cells. The records are approximately homogeneous in the centre area of the cell and have higher values compared to the records in the edges of the cell. The records outside the central area gradually decrease toward the background values. The central areas can be approximated by ellipses with the same centre as the whole cell and the same rotation angle but with smaller

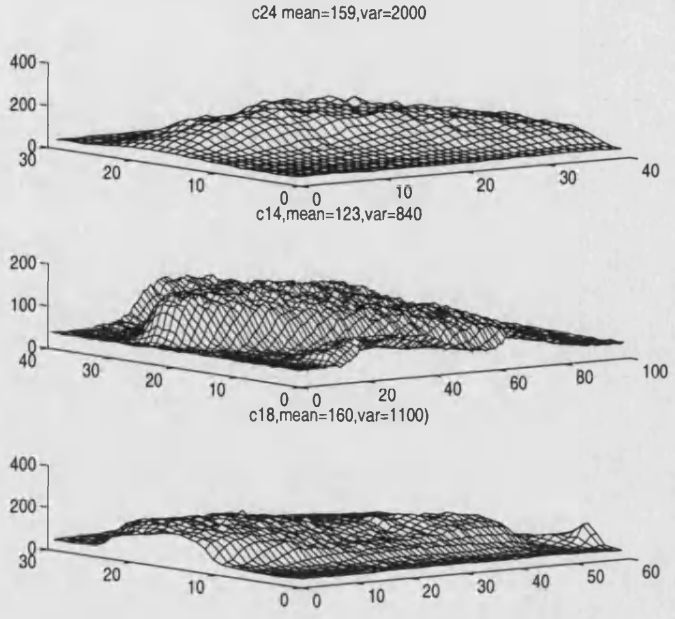

Figure 3-16: 3D image of the records in three different cells

semi-axes. Pawley (1996) explained the phenomena of the heterogeneity of signals in a single cell from the confocal microscopy point of view. He relates the intensity of fluorescence signals to the material in each part of the cell. Higher signals are expected from the nucleus of the cell than the cytoplasm, as it contains higher concentrations of the materials that absorb the fluorescent dye. In the cytoplasm there are changes in the signals depending on the thickness of the cytoplasm area and the angle of imaging of the cell. These dynamic properties will affect the dye distribution and, hence, the reflected signals for different parts of the cells. To reflect this situation, instead of assuming a constant intensity across each cell, we shall allocate different intensity means to different parts of the cell. We define an inner ellipse comprising the central area in which the intensity level is assumed to be constant. The centre of the inner ellipse is the same as that for the cell. The semi-axes for the central ellipse are proportional to the semi-axes of the original ellipse. We define two variables κ_a, κ_b to denote the ratio of the first

semi-axes and the second one respectively. For the pixels lying outside this central area but inside the cell the intensity level will change according to the pixel position. We define, $f_j(\tau_i)$, a function determining the mean intensity level of a record Y_j which falls inside cell i and outside its central area. In considering pixel j lying in cell i, let P be the intercept point of a line connecting the centre of the ellipse l and the centre point j^* of the area covered by pixel j with the border of the inner ellipse (we pick the intercept that is closest to j^*). Let P^* be the intercept of the above line with the border of the cell (again we choose the nearest intercept to j^*). Then $f_j(\tau_i)$ is defined to be

$$\frac{\parallel P^* - j^* \parallel}{\parallel P^* - P \parallel}\lambda_i, \tag{3.44}$$

where $\parallel \cdot \parallel$ denotes the Euclidean distance between two points. Still assuming the records are independent given $X = x$ and normally distributed, the mean of the records Y in this case is

$$\mu_j = \begin{cases} \tau_0 & \text{if pixel } j \text{ is considered as background} \\ \tau_i = \lambda_i + \tau_0 & \text{if the pixel } j \text{ falls in the central area of the cell } i \\ f_j(\tau_i) + \tau_0 & \text{if } j \text{ falls inside } i \text{ but outside the central area.} \end{cases} \tag{3.45}$$

Finally, the variance of the record Y_j is set to be $\sigma_j^2 = \phi_0 + c^*(\mu_j - \mu_0)$.

3.8.2 Modifying the moves

To allow for variability of κ_a, κ_b we assume these to be independent in the reference model with each uniformly distributed over $M_\kappa = (\min_\kappa, 1)$. As we have added extra variables to define an object $x_i = (c_x, c_y, a, b, \theta, \lambda, \kappa_a, \kappa_b)$, then the density for an object defined in equation (3.4) will change to

$$f_X(c_x, c_y, a, b, \theta, \lambda, \kappa_a, \kappa_b) = \frac{1}{\mathcal{A}_L}\frac{|\cos(\theta)| + \frac{1}{\pi}}{3}\frac{z}{2\pi\gamma^2}\exp(-\frac{1}{2\gamma^2}$$
$$(a-\mu_a)^2 + (b-\mu_b)^2)I_{[a\in M_{axis}, b\in M_{axis}]}\frac{1}{\max_\lambda - \min_\lambda}\frac{1}{(1-min_\kappa)^2}, \tag{3.46}$$

The acceptance probability of the birth and death moves will remain unchanged as long as we define the same proposal density as in (3.46) and modify the likelihood term to use μ_j in equation (3.45). For merging two randomly selected objects say x_A, x_B to x_C we use the same techniques and conditions defined in section (3.5.6). We draw $\kappa_{a_C}, \kappa_{b_C}$ uniformly over M_κ. For the reverse split move we also draw four independent random variables uniformly over M_κ. Since we defined these variables independently of the other variables involved in the model, then the Jacobian of the transformation used for the merge move is the product of the Jacobian obtained for the transformation of the $(\kappa_{iz},\ i = a, b,\ , z = A, B, C)$ and the Jacobian obtained from transformation of the other variables. The latter Jacobian remains unchanged as defined in section (3.5.6) and the former is unity. Moreover, since we are using the same proposal densities for the κ as their prior densities, most of the terms involving these variables vanish and the resulting acceptance probability is the same as that of section (3.5.6).

To update κ_a, κ_b of a randomly selected object say, $x_i,\ i = \{1, \ldots, n\}$, we define an extra move $m = 9$ to be chosen with probability $j_9(x)$. In this move we propose κ'_{a_i} and κ'_{b_i} independently and uniformly over $(a-\varrho, a+\varrho)$ and $(b-\varrho, b+\varrho)$ respectively, where ϱ has an arbitrary fixed value. The acceptance rate α for this move is the standard Metropolis rate.

Let $x'_i = (l_i, a_i, b_i, \theta_i, \lambda_i, \kappa'_{a_i}, \kappa'_{a_i})$ be the object after changing κ_a and κ_b, then the acceptance probability for this move for object i is

$$\alpha((\kappa_{a_i}, \kappa_{b_i}), (\kappa'_{a_i}, \kappa'_{b_i})) = \min\{1, R\}, \tag{3.47}$$

where

$$R = \prod_{j \in \Re(x_i)} \frac{\sigma_{(j|x_i)}}{\sigma_{(j|x'_i)}} \exp\left(-\frac{1}{2\sigma^2_{(j|x'_i)}}(y_j - \mu_{(j|x'_i)})^2 + \frac{1}{2\sigma^2_{(j|x_i)}}(y_j - \mu_{(j|x_i)})^2\right).$$

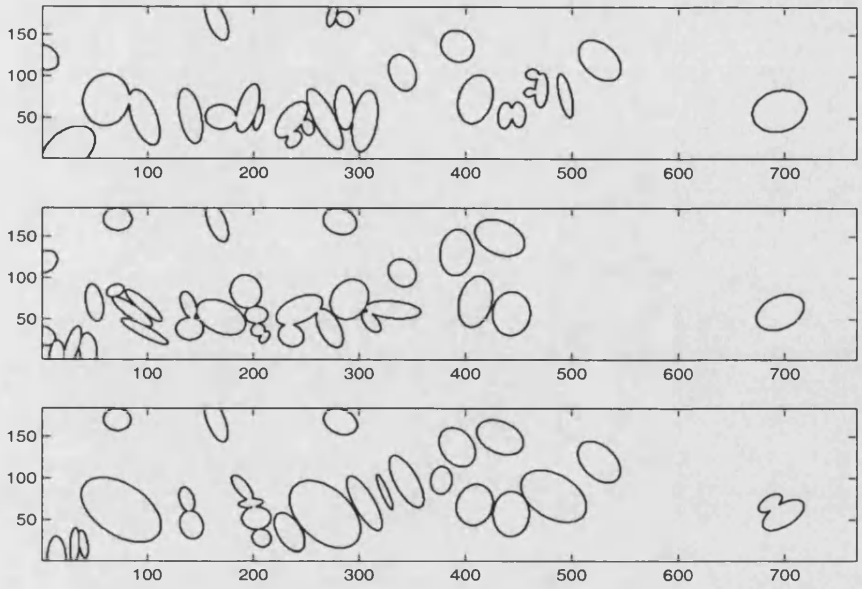

Figure 3-17: The resulting image after applying the κ technique using different runs.

3.8.3 Updating κ_a and κ_b

Assigning a value of 0.7 to $\min_\kappa$ and 0.1 to ϱ, some resulting images after 20000 iterations are shown in figure $(3-17)$. The configurations have positioned suitable cells in some locations of the window. However, some cells which cover large areas have been split into smaller cells. The total number of accepted split moves using this model is higher than the one which uses the original model. Due to the non-homogeneity of pixel records inside the cells with the new mean and the new variance, smaller cells with different intensity level are more representative than big cells. This leads to high acceptance of the split moves.

The output results suggest that the new extension of the model is not managing to guide the simulated samples toward a better fit, even though the modification of the parameter of the likelihood, μ_j is based on exploratory data analysis. This may be due to the random starting values of the mark variables of the cells especially the intensity level λ. We think that with using a good starting point for the chain and using a more representative value for the mark vector as initial values, the model will work better.

3.9 Blurring

The scattering feature of the incoming laser and the outgoing fluorescence in the confocal microscope imply that signals received from focal point will be a measurement of photons from that point along with transmitted fluorescence signals from a surrounding area. Thus, each recorded value in the signal is the sum of contributions from the pixel on which the sensor is focused and from a number of neighbouring pixels. Let Λ_i be the set of labels of pixels in the vicinity of pixel i which affect the value of the record Y_i for pixel i, and let w_j be the positive weight associated with the contribution from pixel j where $j \in \Lambda_i$. We define these weights to sum to one.

Following the assumption of normality and conditional independence of the records given the scene as stated in section (3.1), we express the mean of the blurred records $Y|X$ as

$$E(Y_i|X) = \sum_{j \in \Lambda_i} w_j \lambda_j + \tau_0, \quad i = 1, \ldots S.$$

Here λ_j is the intensity level of the cell covering j and is zero if pixel j is considered a background. The variance of $Y_i|X$, σ_i^2 is associated with this mean as in equation (3.2).

For a uniform blurring kernel, the weights w_j are equal,

$$w_j = \frac{1}{\text{No. of pixels in } \Lambda_i}, \qquad j \in \Lambda_i.$$

Alternatively a Gaussian kernel may be used in which

$$w_j \propto \exp(-\frac{1}{2\upsilon} \parallel l_j - l_i \parallel^2), \qquad j \in \Lambda_i, \tag{3.48}$$

with higher weights for pixels closest to pixel i. Here l_i and l_j are the locations of the pixels i and j respectively and $\parallel l_j - l_i \parallel$ denotes the euclidian distance between l_i and l_j. The value of the variance υ and the size of neighbourhood in the Gaussian

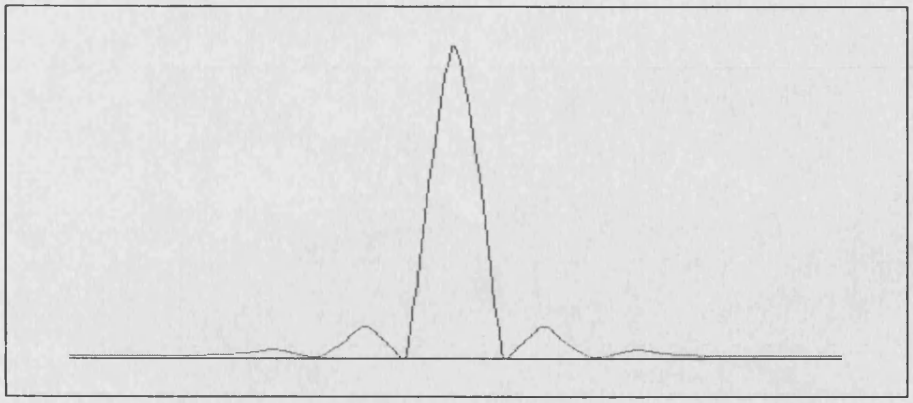

Figure 3-18: The detected intensity as a function of axial position from the focal point in Wilson's experiment measuring the point spread function.

kernel depend on the degree of blurring expected and are a measure of the width of the spread function of measured light from a point source. Experiments to measure the transmitted signals' intensity from the detector plane against different axial position from the focal point have been conducted for a confocal microscope by Wilson (1990). The experiments revealed a fairly narrow peak of the detected signals centred at the focal point and small side-lobes as shown in figure $(3-18)$. This kernel remains approximately constant over the entire image field. Hence, it is reasonable to approximate the point spread function of light by a Gaussian kernel spread function and we use this here. Ongoing developments in the design of the confocal microscopes and the use of different optical systems with, for example, small aperture for the objective lenses, have resulted in improving the signal quality. Due to these improvements, light from a point source illuminates only a small region of the object and the point detector gathers light from only that small area; see Shaw and Rawlins (1991). To assess the blur effect in our model we will define Λ_i to denote the first or the second order neighbourhood. Each pixel i in the two dimensional representation is written as a pair (l,h) where $l \in \{1,2,\ldots,S_x\}$ and $h \in \{1,2,\ldots,S_y\}$.

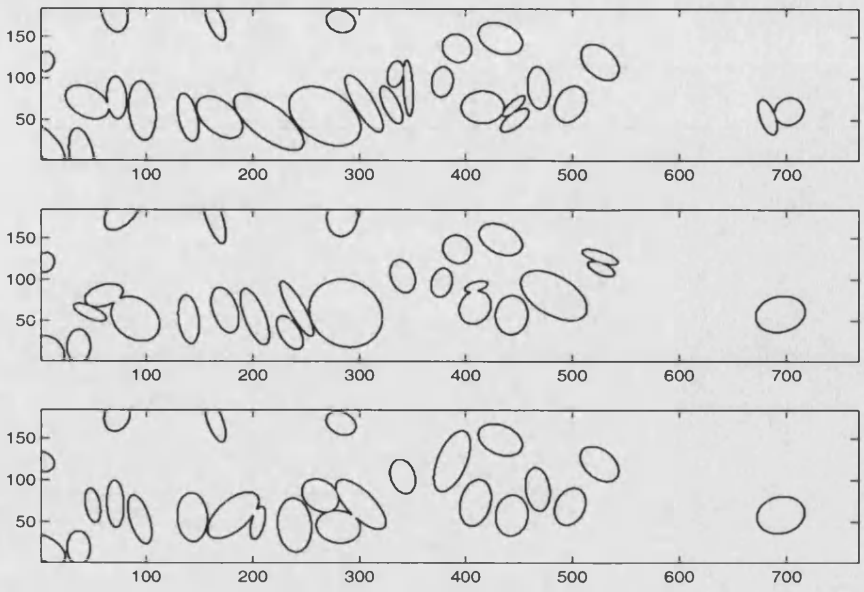

Figure 3-19: Sample images after applying the blur effect evaluated using a first order neighbourhood and $\upsilon = 0.6$.

The first order neighbour structure of the pixel i is the disc region with centre (l, h) and radius one pixel, that is $\Lambda_i = \{(l-1, h), (l, h), (l+1, h), (l, h-1), (l, h+1)\}$. The second order clique is a disc with centre (l, h) and a two pixels radius, $\Lambda_i = \{(l-2, h)(l-1, h), (l, h), (l+1, h), (l+2, h), (l, h-2)(l, h-1), (l, h+1), (l, h+2), (l-1, h-1), (l-1, h+1), (l+1, h-1), (l+1, h+1)\}$. The nth order neighbour structure of pixel i is defined to be a disc centred at (l, h), the centre point of pixel i, and radius equal to n pixels.

Several values for υ the variation of the blur effect were employed to test the blurring effect. Figure $(3-19)$ shows image estimates using a first order neighbourhood and assigning 0.6 to υ. Figures $(3-20)$ and $(3-21)$ represent output samples using $\upsilon = 1.2$ and 6.25 with third and fifth order neighbourhoods respectively. There is no significant difference between $\upsilon = 0.6$ and $\upsilon = 1.2$ but results using $\upsilon = 6.25$ were poor indicating the high quality of the experimental data and the small effect of the blurring. Adding a blurring effect to our model should mainly affect the edges of the

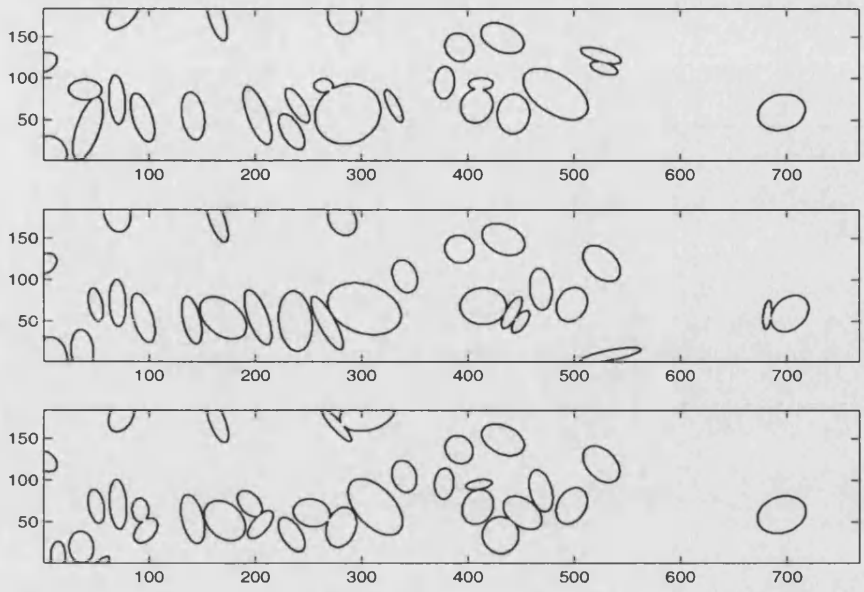

Figure 3-20: The resulting sample images after applying the blur effect evaluated using a third order neighbourhood and $v = 1.2$.

cells and ought to help deal with uncertainty about the borders of the cells. Comparing cells with the same positions and shapes in Figures $(3-19)$ and $(3-20)$, we noticed no significant changes in the cells sizes. Using a bigger neighbourhood for the point spread function will consume more CPU time to run the sampler. Our criteria for comparing different models is through visual inspection of the resulting output samples. Hence, we find no apparent difference between the samples before and after adding a blurring effect. Each cell in the scene covers hundreds of pixels and the small shrink in the size of the cells resulting from using small values of v will not be noticeable. Thus for some general purpose studies where the interest is to locate the cells or to restore the image rather than to give precise point estimates or interval estimates of some mark variables, but the CPU run time is more important we may use the model without blurring.

The examples studied so far indicate that the basic method of defining the posterior model and applying MCMC algorithms for simulating samples from that model works quite well. There are some problems with the slow convergence and getting stuck with a

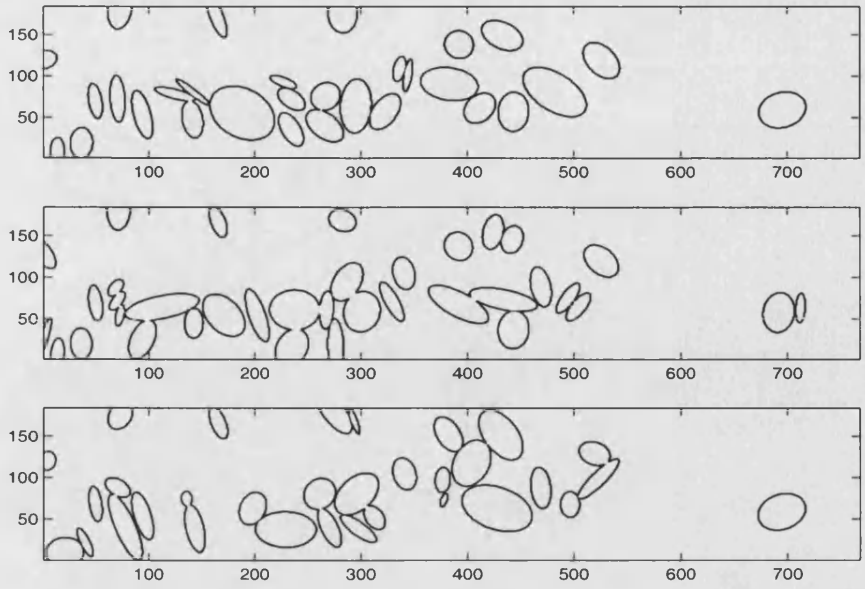

Figure 3-21: The resulting sample images after applying the blur effect evaluated a fifth order neighbourhood and $\upsilon = 6.25$.

few wrongly estimated cells. The feature used for improving the birth move type within the MCMC algorithm has produced promising results and helped in better scheduling of the run time between the different move types. The sensible modification on the likelihood model to agree with the data was not effective using an extreme starting point for the chain, however this model will be used later with another stating state. Finally, models using blurring to reflect the point spread function of the light used in confocal microscopy suggest that applying this modification to the model is recommended if an accurate estimator is needed, but otherwise may be too CPU intensive.

Chapter 4

The performance using good starting states

We have described a model for the two dimensional data set and an MCMC algorithm to simulate samples from that model. In the examples we have studied, the Markov chain takes quite a long time to reach convergence. Moreover, the non-overlapping constraint in the model will negatively affect the speed of the mixing of the chain. Therefore to avoid lengthy run times and to produce samples near the mode of the target distribution, we need to choose a good starting point $X_0 \in \mathbb{E}$, where $\mathbb{E}$ is the space of all possible images. In this chapter we shall describe algorithms for constructing X_0 either by a self-contained approach or with the help of a user. We shall also comment on the simulated set of samples after using the starting state X_0.

4.1 How to construct a good starting state

A general state X_0 can be described as set of non-overlapping objects each defined in the space $\mathbf{U}$. An image, described in this high level manner, has to be estimated from the data observed at the pixel level. Each object covers several connected pixels and parts of pixels as well. In our approach of building up the image model we operate

at the object level. Our posterior model is also a high-level model as it is defined for $x \in \mathbb{E}$. We want to construct high level starting states using the data provided, Y, which is gathered at a low, pixel-based, level. We shall need to process information at different levels, first the pixel level, then the higher object level. We need a device which operates at two consecutive stages. In the first stage we shall operate at a low level and derive structure using some *morphology* concepts. We shall then convert this fine level structure to high level structure using *maximum variance* concept, ready for further processing in the second stage.

4.1.1 Stage I : Mathematical morphology

Mathematical morphology as described by Serra (1988) and Glasbey and Horgan (1995) operates at a low-scale level, interacting with the object under study and modifying its shape. The aim is to produce a new set of objects satisfying some defined criteria more closely than the original objects. Morphology proceeds by applying certain basic operators to binary images. Hence, before we go into details of these operations we need first to transform our grey-scale image into a binary image. We shall use the simple method of *thresholding*. Given a single threshold t, the pixel located at lattice position (i, j), with grey scale, y_{ij}, is allocated to class 1 if $y_{ij} \leq t$, and to class 2 otherwise. In our case, category 1 can be considered as the background region while category 2 is the cell region. The threshold t can be chosen manually. A common value for t is the median or the mean of the image. Let $L_{(xy)}$ be the array representing the discretisation of the image on a two dimensional rectangular lattice. We shall denote by $L^{(b)}$ the resulting binary image which will be used as input for the morphology operators.

Morphological operations use a *structuring element*, B. The structuring element may be a disc of fixed radius, for example, or a rectangle with sides of specified length. A particular pixel in the structuring element is designated as a *reference pixel* and we refer to the structuring element with reference pixel (i, j) as $B_{(i,j)}$. We shall work with

the two basic operations, *erosion* and *dilation*, which are briefly described as follows.

- The erosion operation, $\ominus$: We define the eroded set $L^{(e)}$ of $L^{(b)}$ as the set of (i,j) such that the structural element with reference point (i,j) lies completely within $L^{(b)}$.

$$L^{(b)} \ominus B = \{(i,j) : B_{(i,j)} \subset L^{(b)}\},$$

- The dilation operation, $\oplus$: Using the same element structure, B, we define the dilation of $L^{(b)}$ by B as

$$L^{(b)} \oplus B = \cup_{(i,j)\in L^{(b)}} B'_{(i,j)},$$

  where B' is the rotation of of B through an angle of 180° about (i,j).

Erosion will shrink the original object and eliminate the extreme edges which do not fit into the structuring element. Performing dilation after erosion will enlarge the eroded object and smooth its edges (see figure $(4-1)$). The process of applying both erosion and then dilation is called the *opening* operator $\psi_B(L^{(b)})$ which can be expressed as

$$\psi_B(L^{(b)}) = (L^{(b)} \ominus B) \oplus B' = \cup_{(i,j):B_{(i,j)}\subset L^{(b)}} B_{(i,j)}.$$

In the process of constructing an initial state x_0 for our MCMC algorithm, the stage one scheme can be described as follows

- Threshold the data Y using the median.

- Define a 9×3 pixel-based rectangle as the structuring element, B. The choice of this particular structuring element is to help create components which are roughly elliptical, in addition to smoothing the scene, breaking narrow isthmuses and eliminating small islands.

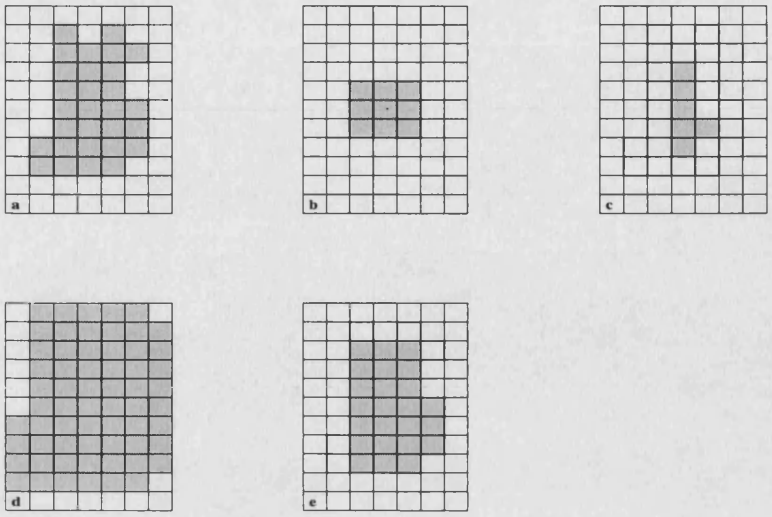

Figure 4-1: Illustration of basic morphology operations: (a) original set, (b) Structuring element B: a square of size 3 pixels (the $\cdot$ in the centre pixel indicate that we have selected this as the reference pixel); (c)The image (a) after erosion with structuring element B, (d) after dilation, (e) using the opening operation.

- Erode the resulting binary image by B.
- Dilate the resulting image using the same element structure B.

Applying the above steps to our 2D optical section image resulted in the image shown in figure $(4-2)$.

4.1.2 Stage II : Maximum variance concepts

In this stage of the construction process, we shall convert the connected pixels of stage I into ellipses. Each object in the image can be looked at as a set of points on the lattice of 2D system with x and y coordinates. These points are the centres of the pixels which make up an object. The set of the connected points has a centre of mass. The set of points is distorted by being stretched in one direction and compressed in the direction perpendicular to the first. That is the set of points is fairly close to an ellipse. Now consider a bivariate normal distribution for (X, Y). This has elliptical

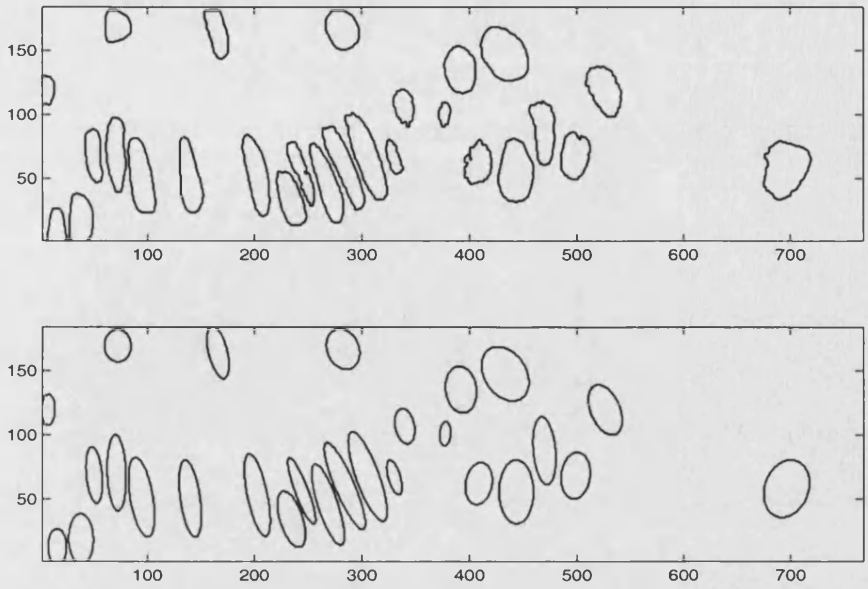

Figure 4-2: The resulting images after applying the morphology operations (top) and then converting the connected pixels into ellipses (bottom).

contours and the set of points at which the density exceeds a set value is an ellipse. We shall treat the set of points at the centres of the pixels in a connected object as a sample from a bivariate normal distribution. Then we shall estimate the parameters of this distribution and create an approximating ellipse from a contour of the bivariate normal density. The bivariate normal distribution with parameters $\mu_x, \mu_y, \sigma_x^2, \sigma_y^2$ and ρ has density

$$f(x,y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{(1-\rho^2)}} \exp(-\frac{1}{2(1-\rho^2)}\{(\frac{x-\mu_x}{\sigma_x})^2 + (\frac{y-\mu_y}{\sigma_y})^2 - 2\rho(\frac{x-\mu_x}{\sigma_x})(\frac{y-\mu_y}{\sigma_y})\}), \qquad (4.1)$$

We estimate the values of the parameters using the sample of (x,y) values from pixel centres in the connected object. For example our estimate of the mean of x, $\hat{\mu}_x$, is taken to be the average value of the $x-$coordinates. The variance σ_x^2 is estimated using

the sample variance s_x^2 and the correlation coefficient ρ is estimated using the formula

$$\frac{\sum_i (x_i - \hat{\mu}_x)(y_i - \hat{\mu}_y)}{n\hat{\sigma}_x\hat{\sigma}_y}.$$

The contour is a cross section of the surface of bivariate normal density made by a plane parallel to the xy-plane, which in essence results from setting the expression in braces in equation (4.1) equal to a positive constant C. The resulting equation,

$$C = \{(\frac{x - \hat{\mu}_x}{\hat{\sigma}_x})^2 + (\frac{y - \hat{\mu}_y}{\hat{\sigma}_y})^2 - 2\hat{\rho}(\frac{x - \hat{\mu}_x}{\hat{\sigma}_x})(\frac{y - \hat{\mu}_y}{\hat{\sigma}_y})\}. \tag{4.2}$$

represents an ellipse with centre $(\hat{\mu}_x, \hat{\mu}_y)$, the centroid of the bivariate population, with rotated major and minor axes.

It can be shown (see Tatsuoka (1971)) that the major and minor semi-axes of the ellipse, defined by equation (4.2), are along the lines passing through the centroid of the bivariate normal density and making the following angle with the positive x-axis and y-axis respectively,

$$\hat{\theta} = \begin{cases} \frac{1}{2}\arctan(\frac{2\hat{\rho}\hat{\sigma}_x\hat{\sigma}_y}{\hat{\sigma}_x^2 - \hat{\sigma}_y^2}) & \hat{\sigma}_x \neq \hat{\sigma}_y \\ 45° & \hat{\sigma}_x = \hat{\sigma}_y. \end{cases}$$

The coordinates of (x, y) with respect to the rotated first and second semi-axes, say x_1 and y_1, are linear combinations of $(x - \mu_x)$ and $(y - \mu_y)$, where

$$x_1 = (x - \hat{\mu}_x)\cos(\hat{\theta}) + (y - \hat{\mu}_y)\sin(\hat{\theta}) + \hat{\mu}_x,$$

and

$$y_1 = -(x - \hat{\mu}_x)\sin(\hat{\theta}) + (y - \hat{\mu}_y)\cos(\hat{\theta}) + \hat{\mu}_y.$$

The length of the major and minor axes, a, b can be determined by calculating the

variances in the directions of the rotated axes x_1 and y_1, that is

$$\hat{\sigma}_{x_1}^2 = \hat{\sigma}_x^2 \cos^2(\hat{\theta}) + \hat{\sigma}_y^2 \sin^2(\hat{\theta}) + 2\hat{\rho}\hat{\sigma}_x\hat{\sigma}_y \cos(\hat{\theta}) \sin(\hat{\theta}),$$

and

$$\hat{\sigma}_{y_1}^2 = \hat{\sigma}_x^2 \sin^2(\hat{\theta}) + \hat{\sigma}_y^2 \cos^2(\hat{\theta}) - 2\hat{\rho}\hat{\sigma}_x\hat{\sigma}_y \cos(\hat{\theta}) \sin(\hat{\theta}),$$

For a given value of C in equation (4.2) the lengths of the major and minor semi-axis are equal proportions of $\max\{\hat{\sigma}_{x_1}, \hat{\sigma}_{y_1}\}$ and $\min\{\hat{\sigma}_{x_1}, \hat{\sigma}_{y_1}\}$ respectively. If we take the length of the major and minor semi-axis to be twice $\max\{\hat{\sigma}_{x_1}, \hat{\sigma}_{y_1}\}$, $\min\{\hat{\sigma}_{x_1}, \hat{\sigma}_{y_1}\}$, then the ellipse will contain a high percentage of the bivariate normal distribution. However, because we want to avoid any overlapping between the resulting ellipses a smaller proportion of 1.50 is used. This scale can be adjusted so that the resulting ellipses do not overlap.

By extracting the required variables for all the different connected pixels we approximate their basic shapes by the ellipses shown in figure $(4-2)$. Finally we need to estimate the mean signals of these ellipses which we take to be the average record value of the pixels contained in each of these ellipses.

Using the resulting figure $(4-2)$ as an initial state, x_0, for our chain, we ran the MCMC algorithm using the posterior image distribution under the original model as the target distribution. Figure $(4-3)$ shows the 20000th and the 50000th sample iterations. The output samples have managed to adjust the existing cells into better positions. The birth of new cells which have not been recognised in x_0 can also been seen in the later images. These new cells did not appear in x_0 because they were not classified as cell media either in the thresholding stage or after applying the morphology techniques. If the cell has low intensity level as has the cell with centre $(52.933, 169.310)$ then it will be classified as background in the thresholding stage. The non recognition of cells at the morphology stage is due to the small size of the sets of the connected pixels; these sets may not be covered by the structuring element, B, and hence, not be

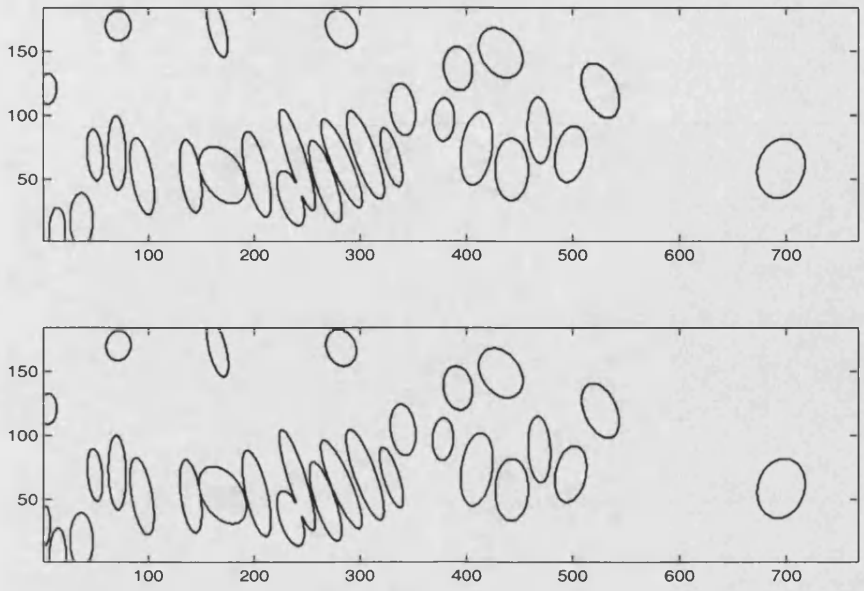

Figure 4-3: Sample images using the representative starting value evaluated at 20000 iterations (top) and 50000 iterations (bottom).

identified as has happened with the cell centred at $(28.685, 1.489)$.

We also ran the MCMC algorithm for the model defined in section (3.8.1) which assumes different intensity levels within each cell after using the starting state generated in this way. The 20000 iteration produced good images as shown in figure $(4-4)$ in contrast to the results obtained starting from an empty configuration. Although the resulting images using this model are good, we shall continue using the original model as it takes less CPU time to run the MCMC sampler.

4.2 Contribution of the knowledge of the expert in implementing MCMC algorithm

In the previous section we used a fully automated, self-contained process in constructing a starting state for the Markov chain. We then used the same model and the same algorithm to implement the MCMC sampler as before. In this section we shall consider

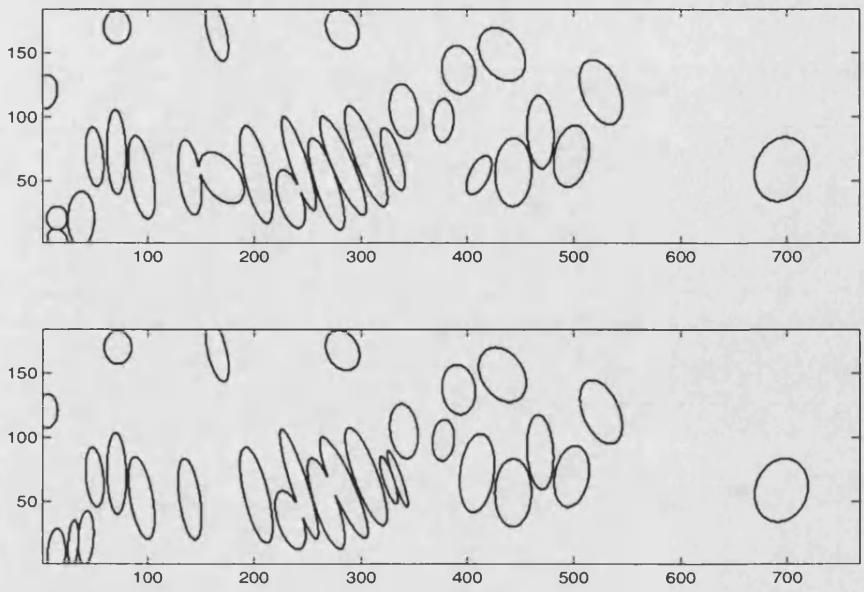

Figure 4-4: Sample images for different runs after applying the κ technique and using a good starting state.

the contribution of the knowledge of an experienced user toward a better and faster convergence of the Markov chain. Suppose the user had provided us with information defining the number of the cells in the scene and had pointed out these cells by specifying a single point inside each cell. This process can be achieved by clicking on the image window and keeping a record of the corresponding coordinates. Different methods of using these points and converting them to ellipses are available. One method is the *seeded region growing* algorithm, Adams and Bischot (1994). The algorithm works for grey-scale images and works on finding the *tessellation* of the images into homogeneous regions based on a given number of seeds. The algorithm operates at a fine pixel-based scale. Essentially it classifies an unlabelled pixel j in L into one of the existing regions based on minimal variance between its record value and the means of different groups. We applied this algorithm to our data and the resulting tessellation was bad. This is because we have large grey-level variations not only in the entire image but also within the cells themselves. Hence, the criteria for labelling the pixels was not suitable for our

data. Another approach was using edge detection algorithms such as Canny algorithm, Canny (1986). Edges are detected by finding discontinuities in the image either by applying some filters to de-convolute the point spread function or using tessellation of the image into homogeneous regions. Both the techniques where not effective for our problem. The latter requires the variation of the records for the entire image to be specified in advance. Applying the general variance extracted from the entire image, which is very high, misclassified the edges of the cells. The white noise variance defined in section (2.1.2) as the variance of the background records ϕ_0^2 is applied using the value 46.5 computed using a sample of background records. This value resulted in a large number of boundaries around each cell. Some of these boundaries were not connected throughout the cells. This is due to the non homogeneity of the records inside the cells and due to the high variation among these records. Using the filtering technique the resulted binary output had a large number of disconnected boundaries.

An alternative approach is to combine the output of section (4.1) with the user information. That is we modify the converted version of the output from the morphology functions which was described in sections (4.1.1) and (4.1.2) to agree with information supplied by the expert. This can be done simply by deleting the non-matching cells and defining new cells if they do not exist, and by splitting an existing cell if it covers more than one point. The algorithm is run using the sensor points clicked by the expert and the resulting ellipses of the morphology and maximum variance concept algorithm. In defining a new cell we shall create an ellipse with centre at the reference (sensor) point, and intensity level taken to be the record associated with the centre point or $\min_\lambda$ whichever is higher. In the absence of any prior information regarding the semi-axes and the angle of rotation, and to avoid overlapping between the objects we shall set the semi-axes to $\min_{axis}$ and $\theta = 0$. The ellipse is deleted if it does not cover any sensor point. If the cell covers more than one reference point then this cell is deleted and replaced by two new cells using the same technique described before. We shall end up with a configuration that agrees with the expert knowledge and partially agrees with

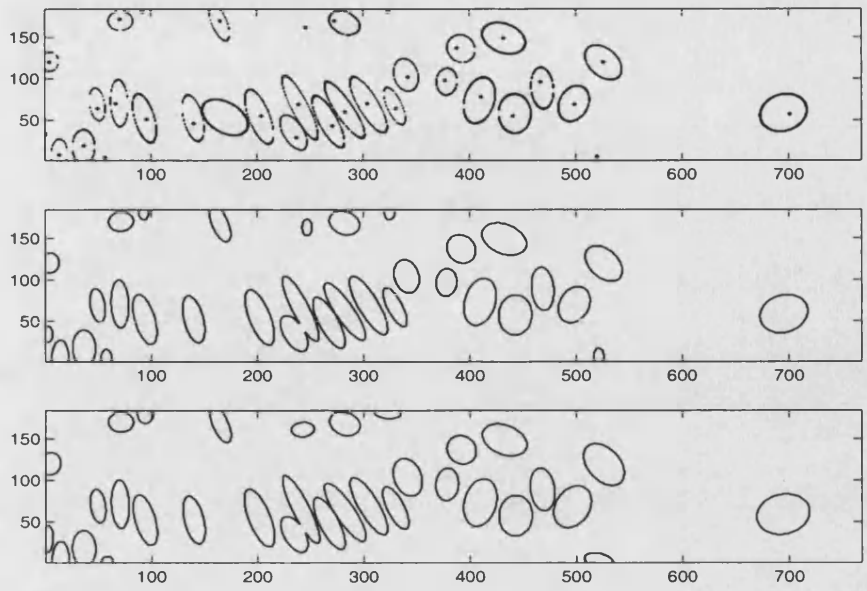

Figure 4-5: Sample image evaluated using the sampler and the reference points (top). The intermediate image which agrees with provided points (middle). The resulting image after applying the conditional sampler using the intermediate image (bottom).

the input image (see figure $(4-5)$). Note here that advantages arise from processing at different levels. In this approach we first start with a coarse level by clicking inside the cells. We then work at a fine level in using the morphology functions, and return to the coarse level again where we convert the morphology output to ellipses and use these as a starting point for the MCMC algorithm.

The confidence level or the reliability of the expert knowledge determines the way to deal with the resulting configuration. If we are not confident about the manual intervention of the user then we can use this output as a starting point for our standard model; the sampler will, eventually after a lengthy run time, produce a result which is invariant of this starting point. On the other hand if we consider the provided information to be totally reliable, then the target model along with the MCMC algorithm should be revised. So, we wish to incorporate expert information and to ensure that all the images created by the MCMC iterations contain precisely the cells specified by the expert and each one always covers the specified point in its interior. As our sampler

during a typical run length for MCMC is thus unable to be invariant of X_0 we shall use the following approach:

We shall start with the reference measure $\Gamma^*(X,N)$ defined on $\cup_n \mathcal{U}^n$, where $\mathcal{U}^n$ is the space of the objects $\{X_1,\ldots,X_n\}$ and define the hard core model with respect to this measure as the prior distribution of X. Then, we shall concentrate on the part of the state space $\mathbb{E}$ with $N=n$ and condition on object i containing a sensor point, say p_i. The relevant contribution from the prior model is

$$\pi(X,N) \propto \beta^n I[\text{no overlap between the objects}] \quad \text{for } N=n.$$

The sub-distribution $\Gamma_n^*(X)$ preserves the number of objects to be n from the Poisson model and comprises the objects $X_1,\ldots,X_N$. The conditional probability measure of x given $N=n$, $\mu(x_1)/\mu(U)\ldots\mu(x_n)/\mu(U)$, is a Lebesgue measure on $\mathbb{R}^{6n}$. With respect to this Lebesgue measure we define for $X_1,\ldots,X_n$ a probability density $\prod_i f_{X_i}$, where f_{X_i} is given in equation (3.6). Substituting for $f_{X_i}, \pi(X,N)$ and for the likelihood function, $\mathcal{L}(Y|(X,N))$ defined in section (2.1.2), and conditioning on the reference points $p_1,\ldots,p_n$ being on the n cells, the posterior probability distribution simplifies to a density on $\mathbb{R}^{6n}$

$$\frac{z_1\beta^n}{((\max_\lambda - \min_\lambda)6\pi\gamma^2)^n}\prod_{i=1}^{n}(|\cos(\theta_i)|+\frac{1}{\pi})\prod_{j=1}^{S}\frac{1}{2\pi\sigma_j^2}\exp(\sum_{i=1}^{n}(-\frac{1}{\sqrt{2\gamma^2}}$$
$$(a_i-\mu_a)^2+(b_i-\mu_b)^2)+\sum_{j=1}^{S}(-\frac{1}{2\sigma_j^2}(y_j-\mu_j)^2))I[a_i\in M_{axis}, b_i\in M_{axis}]$$
$$I[p_1,\ldots,p_n \text{ are in the } n \text{ cells}]I[\text{no overlap}], \qquad (4.3)$$

where z_1 is a normalising constant.

4.2.1 Conditional algorithm

There may be times when the precise objectives of the studies vary. If the number of the objects is defined, then our main aim shall be making inference about the objects'

characteristics. Given that we have the number of the objects in the scene and a reference point p for each cell, we shall not need to use the birth or death moves or split and merge moves. These moves, if accepted, would change the number of the existing cells, and therefore would produce illegal configurations. The only moves to be considered are the moves which adjust the positions or the intensities of the cells conditioned on the reference points.

The performance of this conditional algorithm is evaluated in figure (4−5). To apply the conditional MCMC technique we shall use the random update sampler starting from the state produced in section (4.1) and modified to agree with the expert information. At each iteration we randomly select a cell among the existing cells say i. Then we propose a move type $v, v = 1, \ldots, 4$, where $v = 1$ stands for shift, 2 for resize, 3 for rotation and 4 for updating the intensity of an existing cell. Depending on the move type we propose to adjust the selected object i by changing its mark variable. We shall use the same proposal kernels as defined before for each move. The proposed state will be accepted with the same acceptance rate defined for each of these moves as given in chapter (3) conditioned on having the reference point p of the potential cell inside that cell after modification. If the reference point was not covered by the proposed cell then we shall reject the move. For example suppose the shift move is selected, then we propose to update x_i to x_i' by proposing new location l_i'. The new x_i' will be accepted with probability $\alpha(l_i, l_i')I[p_i \in x_i']$, where $\alpha(l_i, l_i')$ is as defined in equation (3.26). Similar concepts are applied for resize and rotation moves. For the move that updates the intensity level this condition need not be checked as this move will not alter the position of the cell and, hence, we use the same acceptance rate given by equation (3.30).

If we are going to use the conditional algorithm, then a higher degree of coordination can be achieved by using the MCMC sampler along with the fully automated starting point technique and with user intervention. That is instead of directly comparing the fully automated starting state X_0 with the user's input points, we allow the

unconditional MCMC sampler to run for a long time. Then we shall ask the user for his or her intervention and adjust the resulting MCMC sample to coincide with this information. Finally we run the conditional sampler for another reasonable time to obtain the sample output. The multi-steps series can be illustrated as follows:

1. Construct an automated starting point.

2. Run MCMC sampler using the unconditional algorithm.

3. Allow for manual intervention.

4. Adjust the sample output.

5. Run the conditional MCMC sampler using the new initial state.

The advantage of this technique is that, since we are dealing with highly reliable expert knowledge, we should employ some statistical tools and assumptions to help the user in making decisions regarding the needed information. Also, if most cells have already been identified, the expert can do what is needed quickly.

To test the performance of the conditional algorithm a run of 20000 iterations was carried out using the intermediate image obtained from the previous unconditional run, modified after applying some manual intervention. The starting state is displayed in figure $(4-5)$. The output configuration corresponding to the 20000th iteration is shown in the bottom part of figure $(4-5)$. Visual inspection of the output samples in comparison to the input image indicates that the Markov chain managed to converge to the modal region of the posterior image distribution. Along with visual inspection of the output samples, we use another criterion for testing the convergence. We use the traces of the likelihood, the prior and the posterior density to provide additional diagnosis for the convergence of the Markov chain to the posterior model. The run was extended to 100000 iterations and values of the indicators were recorded throughout the run time and are displayed in figure $(4-6)$. All the move types defined for the conditional algorithm preserve the dimension of the parameter space, thus it is feasible

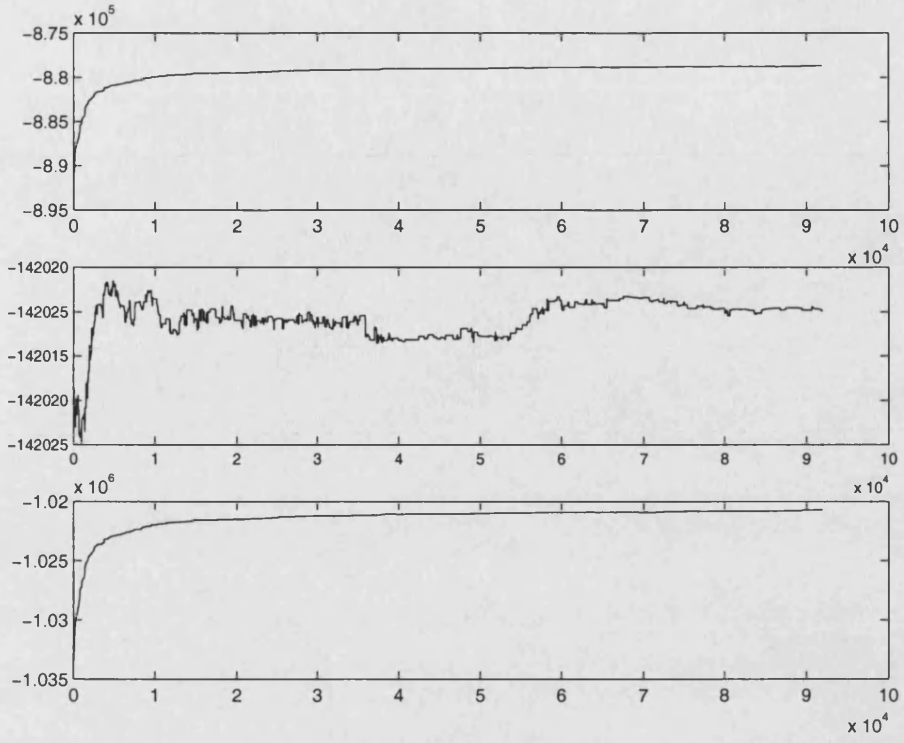

Figure 4-6: A trace of the log likelihood (top), log prior density(middle) and log posterior density(bottom) recorded through out a run time using a representative starting state and a conditional algorithm. The vertical axis represents the observed value in logarithmic scale while the horizontal axis is the number of iterations.

and convenient to trace the prior and hence the posterior density for this sampler. The posterior density is proportional to the product of the likelihood and the prior density. For this particular subspace, it is apparent that the likelihood dominates the prior model in the calculation of the posterior image density. This is because the data set is very informative. For the 100000 iterations the log likelihood ranges from -8.9×10^5 to -8.8×10^5, while the range of the prior density in the logarithmic scale is just 10. The prior density changed rapidly during the first iterations but more steadily as the number of iterations increased. These findings demonstrate the strong influence of the likelihood on determining the shape of the posterior image distribution.

Throughout the run time, the likelihood and hence the posterior density increased in different rates. At the early stages of the run, the likelihood and the posterior

increased rapidly, after that they increased more slowly. It is debatable whether the values are still increasing at the end of the 100000 iterations. Although it is difficult to observe the rate of the increase in the late stages of the run by the naked eye, small changes in the figures correspond to significant changes in the values of the likelihood and the posterior density as we are considering a logarithmic transformation which has been scaled by 10^5 and 10^6 for the likelihood and the posterior density respectively.

Comparing the likelihood results with those obtained using an empty configuration starting state and an unconditional algorithm (see figure $(3-10)$), one can obviously realise that the values of the likelihood have changed significantly. Apparently a good starting point for the chain is a crucial need for convergence within a reasonable amount of time. Using a representative starting state, we managed to overcome the difficulties of the mixing problems resulted from the overlapping constraint and placed the chain in a region near the mode of the target distribution. In the first 4000 (to 10000) iterations the sampler had adjusted itself toward a better position, once it is in the modal region the sampler moves slowly toward the global mode. Hence, a burn-in time is still needed even if we start with a good starting state.

Chapter 5

Image interpretation

Image interpretation is the task of extracting suitable representations of the image content and devising suitable decision making schemes which interpret the content in terms of their representations.

In the previous chapter we have tested the performance of our model and fitting algorithm when using a good starting state. The results were promising and we are more confident that the output result of this method is near the mode of the target distribution. Therefore we can now use the output samples to extract some fine details of the image.

In this chapter we shall classify the cells in the scene into two types and extract some aspects of interest for the biologists for these two types of cells.

5.1 Types of cartilage cells

The data we present in this thesis are from an area of cartilage growth. The zonal subdivision of most of the cartilage tissue can be categorised into four different zones, Woessner *et al.* (1993): The tangential layer, middle layer, deep layer and the calcified cartilage layer. The middle and the lower layers occupy around 80% – 90% of the centre of the mass of a cartilage tissue; they play a big role in the differentiation of

the cellular tissue. These two layers can be distinguished by the content of their cells. The middle layer contains more rounded cells, while the deep layer will contain more elliptical shape cells arranged in columns. The deep zone can be recognised in the area covered by the 2D region $(40, 100) \times (50, 300)$ in the image represented in figure $(3-1)$. The middle layer can be distinguished in the input image to be the 2D region $(40, 160) \times (350, 600)$. The remaining two layers are regions in the two edges of the image specimen. We are mainly interested in the cells in stages two and three, because the information about the two main layers provides the possibility of recognising stages of differentiation and maturity. Burkitt *et al.* (1993) have defined the differentiation and the maturity as follows: "The differentiation of the of-stellate-shaped deep layer cells forms rounded cartilage cells called *chondroblasts.* Sequence division gives rise to aggregation of closely packed chondroblasts which then grow. These cells are then separated from one another because of the secretion of an ex-cellular material. Mature cartilage cells known as *chondrocytes* maintain the integrity of the cartilage matrix. Towards the periphery of the cartilage, the chondroblasts merge with the surrounding supporting tissue to form zone one and zone four". The proportion of area covered by each zone in the tissue depends on the age, condition of the tissue and the amount of physical pressure on that part of cartilage tissue. In imaging using the microscope the plane of vision chosen affects the appearance of these zones as well.

We shall be referring to the chondroplasts cells in the deep layer as type I cells, and using the label type II for cells in zone two. Our only criterion for distinguishing the two types of cells is their shape. Although both shapes can be approximated as ellipses, the eccentricity of type I cells is larger than it is in type II cells. The eccentricity of an ellipse is a measure of the variation of its semi-axes and is defined as

$$ec = \frac{\sqrt{a^2 - b^2}}{a}.$$

The prior knowledge elicited is that the ratio of the major axis to the minor axis tends

to be around two in type I cells and approximately one in type II cells. To reflect this situation we shall assume that the specimen under study contains just the two types of cells and that these are randomly located in the window L. The main difference in these two types of cells are the distributions of their semi-axes. No other constraints shall be applied in the model. The type of the cells, which we shall represent as tp is an identifying variable of a cell, hence, it should be added to the mark vector m. The mark vector defined in section (3.2) is then extended to be $(tp, a, b, \theta, \lambda)$ and is defined on $M = \{1,2\} \times M_{axis} \times M_{axis} \times M_\theta \times M_\lambda$. With the lack of prior knowledge regarding the relative frequency of the two type of cells, in the reference measure, we shall assign an equal probabilities for the two types,

$$\pi_{TP}(tp) = \frac{1}{2}, \quad tp = 1, 2.$$

To model the variability of the semi-axes of the cells we shall define an appropriate conditional density for each type. We assume that, given the type of the cell, the two semi-axes are independent and have a bivariate normal distribution. The mean and the variance for these two distributions are different, each reflecting the properties of its type. For type I cells, the two semi-axes have means 25.0 and 9.0 respectively and equal variance $\gamma^2 = 2.5^2$. Thus the covariance matrix for the bivariate normal distribution is

$$\Sigma = \begin{bmatrix} 2.5^2 & 0 \\ 0 & 2.5^2 \end{bmatrix}.$$

For type II cells the mean vector is $(19.5, 17.5)$ with the same covariance matrix as defined for type I cells. The values for the mean vector and the variances are extracted from output samples of the MCMC algorithm when using a representative starting point.

This extension of the mark vector will be handled in the reference measure. Each object is now defined on $\mathcal{U} \subset \{1,2\} \times \mathbb{R}^6$, in this space we define a density f_{X_i} with

respect to a measure which is a product of discrete measure on $\{1,2\}$ and a Lebesgue measure on $L \times M_{axis} \times M_{axis} \times M_\theta \times M_\lambda$ where

$$f_{X_i}(c_x, c_y, tp, a, b, \theta, \lambda) = \pi_l(c_x, c_y)\pi_{TP}(tp)\pi_{(a,b)|tp}(a, b)\pi_\theta(\theta)\pi_\lambda(\lambda). \tag{5.1}$$

Here $\pi_l(c_x, c_y), \pi_\theta, \pi_\lambda$ are as defined in section (3.3). For the states with $N = n$ objects, the posterior probability distribution is a sub-density with respect to Lebesgue measure and it is proportional to

$$\frac{\beta^n}{(\max_\lambda - \min_\lambda)^n} \prod_{i=1}^{n} [\frac{|\cos(\theta_i)| + \frac{1}{\pi}}{3} \prod_{tp=1}^{2} \frac{I[TP_i = tp] z_{tp}}{2 \times 2\pi\gamma^2} \exp\{-\frac{1}{2\gamma^2}((a_i - \mu_{a_{TP}})^2 +$$

$$(b_i - \mu_{b_{TP}})^2)\} I_{[a_i \in M_{axis}, b_i \in M_{axis}]}] \prod_{j=1}^{S} \frac{1}{2\pi\sigma_j^2} \exp(-\frac{(y_j - \mu_j)^2}{2\sigma_j^2}) I[\text{no overlap}]. \tag{5.2}$$

Here z_{tp} is a normalising constant for the conditional distribution of (a, b) given the type of the cell is tp.

5.1.1 Relabelling move

To implement MCMC algorithm we shall add a relabelling move to the previously defined moves. Suppose our current configuration is $x = \{x_1, \ldots, x_n\}$ where $x \in \mathcal{U}^n$ and the relabelling move is selected. We randomly select an object from $\{x_1, \ldots, x_n\}$, say x_i. Conditional on all variables of x_i we propose to change its type from tp to tp'. The proposal density q is a degenerate unit point mass. That is the proposed change is deterministic, given we have selected to make a move of this type. Let x_i' be the object after changing its type, then the acceptance probability of this move is

$$\alpha(tp_i, tp_i') = \min\{1, \exp(-\frac{1}{2 \times 2.5^2}[(a_i - \mu_{a_{tp_i'}})^2 + (b_i - \mu_{b_{tp_i'}})^2$$

$$-(a_i - \mu_{a_{tp_i}})^2 - (b_i - \mu_{b_{tp_i}})^2]\} \tag{5.3}$$

5.1.2 Changes in implementing other moves

Some changes in implementing the other move types will occur due to the change of the model. If we are going to use the unconditional algorithm that is with an unknown number of cells, we shall allow the birth, death, split and merge moves.

In the birth move we propose a new cell by first proposing its centre, type, angle of rotation and intensity using their prior marginal densities. We shall propose the type uniformly on the set $\{1, 2\}$ and conditional on the proposed type when we draw the semi-axes, a, b randomly using their conditional distributions. Although the procedures for the birth move has changed from that defined in section (3.5.1), the acceptance rate remains the same as given in equation (3.21).

For the resize move there will be no changes in the procedure of implementing the move. However, the changes will occur in choosing the correct parameters for the semi-axes density conditioned on the type of the randomly selected object to update. Equation (3.35) will hold, replacing μ_a and μ_b by $\mu_{a_{tp_i}}$ and $\mu_{b_{tp_i}}$ respectively.

For the split and merge moves two different options may be considered:

- We may constrain merges to be between two identically labelled cells. Hence, we would then propose to split cells into two cells having the same type as the original split cell.

- Alternatively we may propose an unconstrained type in merging cells, and accordingly split any type of cell into random type cells.

The second pair of move types is rather more general than the first case. We shall use this alternative as there is uncertainty about the type of the existing cells. We propose to merge two randomly chosen cells if they satisfy the conditions stated in section (3.5.6). The label of the new cell shall be randomly drawn using the equal probabilities of the two types of cells. As we need bijection between this move and the reverse split move to maintain the detailed balance condition, we shall randomly assign types to the resulting split cells. This is valid as the probability densities of

the semi-axes for both the types are valid for the semi-axes space, M_{axis}. Suppose that the current state is $x = \{x_1, \ldots, x_{n-1}, x_n, x_{n+1}\}$ and we propose to combine two randomly selected cells say $\{x_n, x_{n+1}\}$ to x_n^* which if accepted will take the current state to $x' = \{x_1, \ldots, x_{n-1}, x_n^*\}$. The first term involved in calculating the acceptance rate for the merge move is equal to

$$\frac{12 z_1^{-1} \pi \gamma^2 (|\cos(\theta_n^*)| + \frac{1}{\pi})(\max_\lambda - \min_\lambda)2}{\beta(|\cos(\theta_n)| + \frac{1}{\pi})(|\cos(\theta_{n+1})| + \frac{1}{\pi})} \exp[-\frac{1}{2\gamma^2}((a_n^* - \mu_{a_{tp_n^*}})^2 + (b_n^* - \mu_{b_{tp_n^*}})^2 - (a_n - \mu_{a_{tp_n}})^2 - (b_n - \mu_{b_{tp_n}})^2 - (a_{n+1} - \mu_{a_{tp_{n+1}}})^2 - (b_{n+1} - \mu_{b_{tp_{n+1}}})^2] I[\text{no overlap}] I[\min_{axis} \leq a_n^*, a_n, a_{n+1}, b_n^*, b_n, b_{n+1} \leq \max_{axis}]. \quad (5.4)$$

Here z_1 is the ratio of the normalising constants relative to the densities of the semi-axes. As $q(tp_n)q(tb_{n+1})/q(tb_n^*) = 1/2$, the second term of the acceptance rate of the merge move, $j_{split}(x')q(u)(n+1)/2j_{merge}(x)$, is half the one obtained in section (3.5.6) and displayed in equation (3.39). The other terms involved in the calculation of the acceptance rate remain the same as in section (3.5.6). The acceptance probability for the reverse split move follows from the acceptance rate for the merge move.

5.2 Extracting the important properties of the cells

Properties related to the types of the cells such as the ratio of the numbers of the two types of cells, the eccentricity of each type, their circumferences and their sizes are the most informative criteria for the biologists to understand the basic process in the cartilage tissue. These provide the basis for understanding the *proliferation* and *differentiation* of various cellular components of cartilage tissues and elsewhere; for more biological details see Noda (1993). To estimate this information we shall need the distribution of the properties under the posterior density. We shall extract this information from the Markov chain using a good starting state.

In order to trace the individual variables involved in defining the cells, we shall fix the number of the cells in the scene and fix their approximate location, i.e., we shall

use the conditional algorithm defined in section (4.2.1). Using this technique helps us keep track of the cells and implements moves which are relative to our interest. We run the chain for sufficient time to ensure that the chain reaches the global modal region of the model and that the chain mixes well enough to calculate the variation associated with each variable. We consider 10% of the total run time as a burn-in period. We use a random scheme update; at each iteration we propose a random move to update the current configuration. The iteration is completed after implementing the proposed move regardless of the acceptance or rejection of the proposed state. The consecutive output samples are highly correlated. Therefore we shall allow a reasonable time, fixed in advance, between the collected samples to allow a good mixing of the sub-sampled chain.

For each cell in the sample we extract the variables of interest such as eccentricity, area and intensity. The user may be interested in individual aspects of some specified cells or he may want the general aspects of each type of cells. For the latter case the extracted information from each cell is then averaged over all the cells of the same type and records are made using these averages and this is repeated throughout the collected samples. Using the resulting values we obtain point estimates and associated interval estimates of the required variables. The mean of each of these variables provides a point estimate for our variables. For the interval estimate we need to compute the variance of these records as well. The records are correlated and, therefore, the sample variance may not be accurate if the sample size used is not sufficient. However for efficient subsamples of a very large number of samples we do need not to worry about the correlation. This can be checked using the estimated *integrated autocorrelation time*, $\mathcal{T}$. We only do not need to worry if $\mathcal{T} = 1$, otherwise we use $(\sigma^2\mathcal{T})/n$ in interval estimates, where σ^2 is the variance of the variable needed to estimate and n is the sample size. Green and Han (1992) and Geyer (1991) describe some approaches to estimate $\mathcal{T}$.

Figure $(5-1)$ is a sample after 50000 iterations classifying the type of each cell in

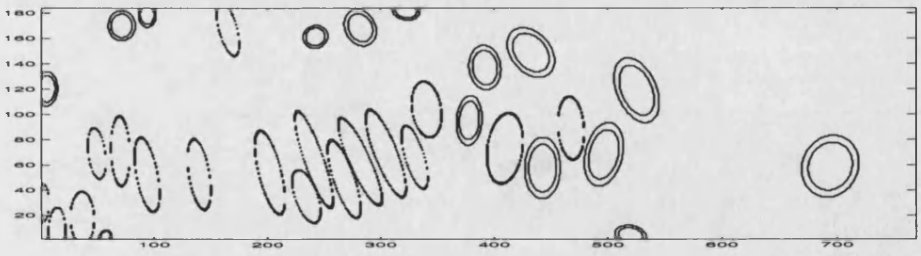

Figure 5-1: A resulting sample image classifying the cells into two types. Type I are the cells with a single-line borders and type II cells are outlined with double-line borders.

the configuration. The classification of the type of each cell was not according to the final label of that cell but rather on the proportion of time the cell was labelled as type tp, $tp = 1, 2$, during the attempts to update the label of this cell in the run time. That is we use the marginal posterior mode as the estimator of type. In figure $(5-1)$ the cells with a single border are type I and type II cells are those with a double border line. Figure $(5-2)$ shows the histogram of the required aspects of each type of cell using the samples evaluated every 250 iterations after neglecting the first 5000 iterations. The discrete nature of the histogram is in part due to the difficulties in samplers mixing. The proportion of type I in the total number of cells was 0.57. Figures $(5-3)$,$(5-4)$ are histograms of the required variables for two individual cells centred (88.594,468.133) and (69.336,497.755) respectively. Cell (1) which its centre located at (88.594,497.755) is considered of type I with probability 0.89, while the probability of classifying cell (2) with centre (69.336,497.755) as type I is 0.21. The estimated mean, the standard error and the $\mathcal{T}$ are computed and listed in table (5.1). In spite of the differences between the estimated values of the variables using the three different observations: a typical cell, cell(1) and cell(2), one can conclude that the size of the area of type I is larger than that of type II. It is worth mentioning that the values of the estimator of a typical cell of a particular type are calculated by considering all the cells classified to have that label. We extract the variable concerned from each cell in that category and then

Variable	Type	Mean	S.E.	$\mathcal{T}$
Typical cell				
Eccentricity	I	0.641	0.02	2.189
Area	I	826.388	26.87	3.58
Intensity	I	78.448	2.22	2.265
Eccentricity	II	0.252	0.02	2.482
Area	II	654.136	43.82	3.914
Intensity	II	74.455	2.47	2.101
Cell(1)				
Eccentricity	I	0.511	0.27	5.982
Area	I	1112.221	231.62	4.402
Intensity	I	54.410	10.75	3.363
Eccentricity	II	0.364	0.02	4.915
Area	II	1013.432	6.68	4.915
Intensity	II	57.724	1.27	1.638
Cell(2)				
Eccentricity	I	0.572	0.07	4.511
Area	I	871.647	112.62	2.345
Intensity	I	67.780	4.57	8.280
Eccentricity	II	0.375	0.01	2.351
Area	II	729.099	270.31	2.733
Intensity	II	70.615	5.51	2.304

Table 5.1: The estimated mean, standard error and integrated autocorrelation time for the different variables for a typical cell, cell(1) with centre (88.59,468.13) and cell(2) with centre (69.34,497.76).

average over those values.

5.3 Sensitivity analysis

The robustness of the our estimates to the assumptions in our model is a common concern of the biologists. In section (3.8) we addressed some aspects of sensitivity analysis, by examining several choices for the hyper-parameters involved in the likelihood model. In this section we shall concentrate on the sensitivity of the inferences to the assumptions in the prior distribution. A common technique to evaluate the robustness of our model, is to use different distributional assumptions and then compare the results with

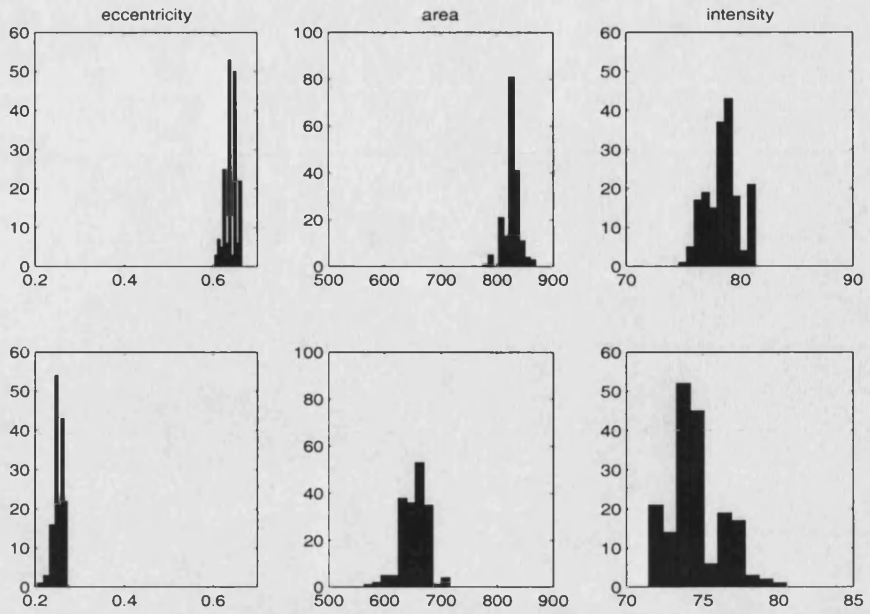

Figure 5-2: Histogram of eccentricity, area and intensity for type I cells (top) and type II cells (bottom).

those obtained previously. There are several reasonable models that can be used for the mark variables instead of the specified ones.

We had defined a uniform distribution over M_λ for the intensity level of the cells to reflect the absence of any more specific prior knowledge. An alternative is to use a Gamma distribution. The values of the mean and the variance of the intensity level λ can be extracted from samples of the cells as plotted in figure $(3-3)$, or from output configurations of previous MCMC samplers. Two different Gamma distributions are employed, one with mean 70.0 and variance to 876.0. The second one has a mean of 100.0 and a variance of 950.0. These two choices of the parameters of the Gamma distribution have different effects: the first, reflects mostly the cells with low intensity level, while the second, favours high intensity cells.

The semi-axes of the cells a, b are assumed to be independent and having a bivariate normal distribution with mean vector $(25.0, 9.0)$ for type I cells and $(19.5, 17.5)$ for type

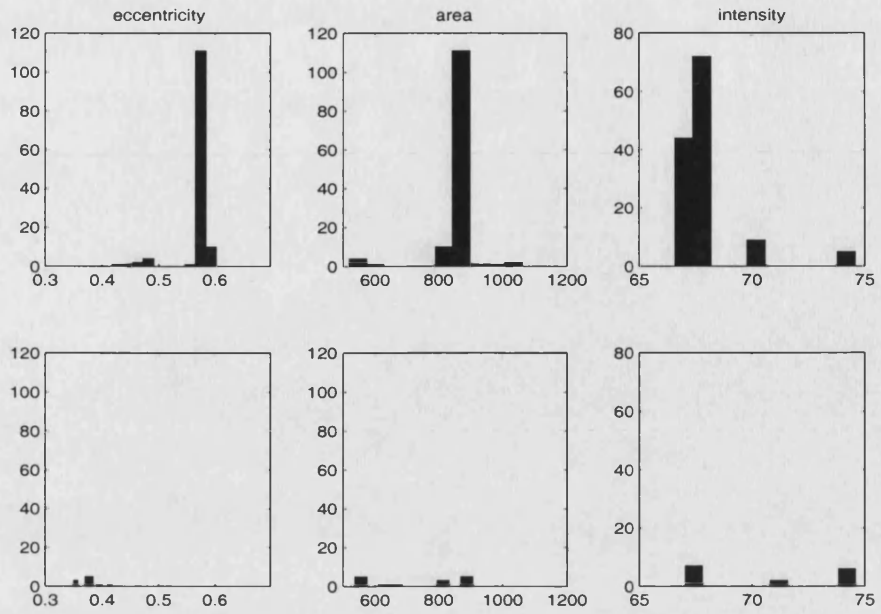

Figure 5-3: Histogram of eccentricity, size and intensity for the cell with centre point (88.594,468.133) when it was classified as type I cell (top) or type II cell (bottom).

II cells and the same covariance matrix Σ, where

$$\Sigma = \begin{bmatrix} 2.5^2 & 0 \\ 0 & 2.5^2 \end{bmatrix}.$$

We shall still assume that the semi-axes are normally distributed but shift the values of their means by 5.0 and -5.0 so to induce changes in the size of the cells of the configurations.

The density of the angle of rotation, θ, is defined such that it reflects the positions of the cells of the input image. We defined

$$\pi_\theta(\theta) = \frac{|\cos(\theta)| + \frac{1}{\pi}}{3},\ 0 \leq \theta \leq \pi.$$

As an alternative, we shall use a uniform distribution which has an equal mass over the entire region of $M_\theta = (0, \pi)$. This is particularly important as it relaxes the previous

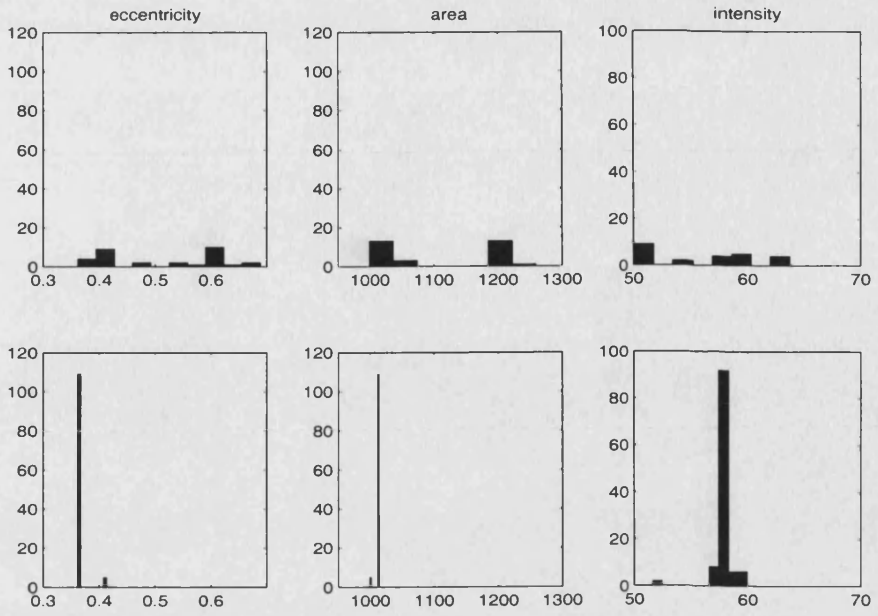

Figure 5-4: Histogram of eccentricity, size and intensity for the cell with centre point (69.336,497.755) when it was classified as type I cell(top) or type II cell(bottom).

distribution and uses a flatter one instead.

Each of these changes is implemented separately and runs of the conditional algorithm were carried out. Samples of the posterior density were then used to evaluate the point estimates (empirical average), integrated auto correlation time $\mathcal{T}$ and estimates of the standard error of the point estimate for the geometric properties of the cells; eccentricity, area and intensity of the two types of the cells. The results are listed in table (5.2),(5.3) and (5.4). The results offer evidence that the different distributional assumptions of the priors of λ, θ and the semi-axes a, b give almost the same results when applied to our model. These unsignificant changes in the values of our estimators indicate that the results are not sensitive to the prior assumptions of λ, a, b and θ. So we can conclude that our posterior model is robust to the different choices of the prior distribution.

Variable	Type	Mean	S.E.	$\mathcal{T}$
mean = 70.0, variance = 876.0				
Eccentricity	I	0.640	0.021	2.29
Area	I	826.893	27.42	3.63
Intensity	I	78.416	2.238	2.34
Eccentricity	II	0.253	0.018	2.57
Area	II	650.439	44.53	3.97
Intensity	II	74.455	2.487	2.15
mean = 100.0, variance = 950.0				
Eccentricity	I	0.641	0.021	2.18
Area	I	826.388	26.871	3.58
Intensity	I	78.449	2.214	2.26
Eccentricity	II	0.252	0.017	2.48
Area	II	654.136	43.795	3.91
Intensity	II	74.461	2.458	2.08

Table 5.2: The estimated mean, standard error and integrated autocorrelation time for the different variables for a typical cell when applying a Gamma distribution for the intensity level of the cells. The top part corresponds to Gamma distribution with mean 70.0 and variance 876.0 and the second part corresponds Gamma distribution with mean 100.0 and variance 950.0.

Variable	Type	Mean	S.E.	$\mathcal{T}$
Shifting of the mean vector of a and b by 5.0				
Eccentricity	I	0.590	0.012	1.44
Area	I	822.546	22.44	1.69
Intensity	I	75.445	1.175	1.98
Eccentricity	II	0.186	0.018	2.61
Area	II	595.780	37.24	1.44
Intensity	II	79.307	2.238	1.26
Shifting of the mean vector of a and b by -5.0				
Eccentricity	I	0.673	0.013	2.93
Area	I	822.671	42.706	5.54
Intensity	I	81.744	2.135	3.68
Eccentricity	II	0.269	0.013	1.75
Area	II	685.271	38.678	4.22
Intensity	II	71.686	1.489	3.84

Table 5.3: The estimated mean, standard error and integrated autocorrelation time for the different variables for a typical cell when shifting the mean vector of the bivariate normal distribution of the semi-axes a and b by 5.0 (top) and -5.0 (bottom).

Variable	Type	Mean	S.E.	$\mathcal{T}$
Eccentricity	I	0.641	0.022	2.27
Area	I	824.893	28.425	3.61
Intensity	I	78.436	2.240	2.36
Eccentricity	II	0.255	0.037	2.47
Area	II	650.430	418.02	3.88
Intensity	II	75.488	2.485	2.12

Table 5.4: The estimated mean, standard error and integrated autocorrelation time for the different variables for a typical cell when defining a uniform distribution for, θ, the angle of rotation.

Chapter 6

Generalising the problem from the 2D plane to 3D space

6.1 Introduction

The advantage of confocal microscopy is that it is possible to image the 3D specimen without physical sectioning. Light from a laser is focused by an objective lens onto a point in the specimen, the emitted light is then refocused by a collector lens onto a detector. By changing the focal point within the horizontal plane to cover the entire specimen, a 2D optical image is obtained. The three dimensional representation of the specimen is acquired by scanning consecutive sections with a change of focal depth between each section (optical serial sectioning). Our experimental data are a 3D image for an area of cartilage growth. In the previous chapters we dealt with one of the 2D optical sections and built up a statistical model for the 2D image. In this chapter we shall generalise this model to 3D images

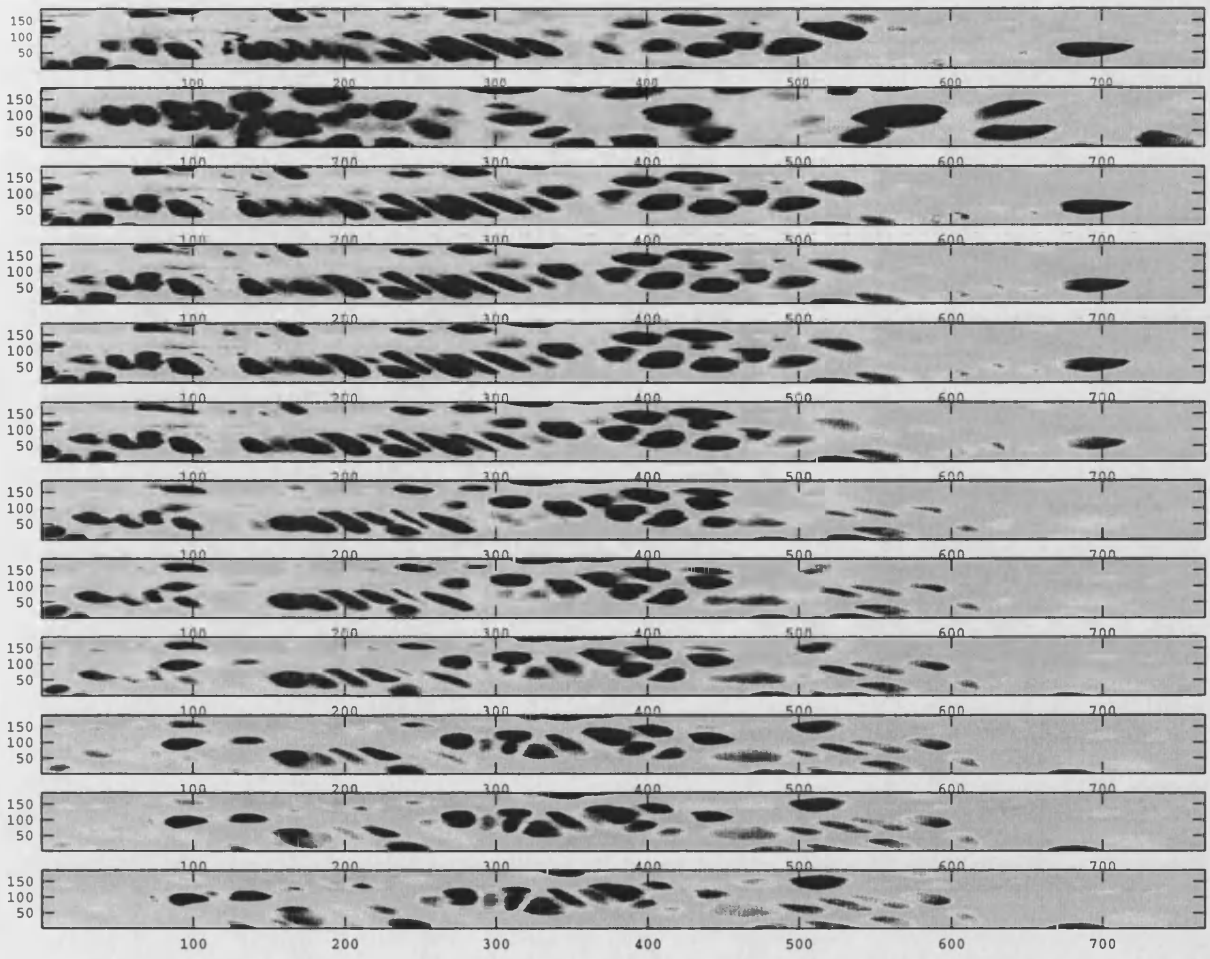

Figure 6-1: The 3D input data represented as consecutive 2D optical sections.

6.2 The likelihood function

The data are presented as $Y_{ijk}, i = 1, \ldots, S_x, j = 1, \ldots, S_y, k = 1, \ldots, S_z$, where S_x, S_y, S_z stand for number of rows, columns and stacks along the $z-$axis respectively. In the grid we have $S_x = 184$, $S_y = 768$ and $S_z = 12$. In general the width of rows and columns in the 2D sections is different from that used for the depth distance. This scale is controlled by the expert user and is chosen according to the specimen under study. The accuracy of the 3D imaging extends for a defined distance through the data

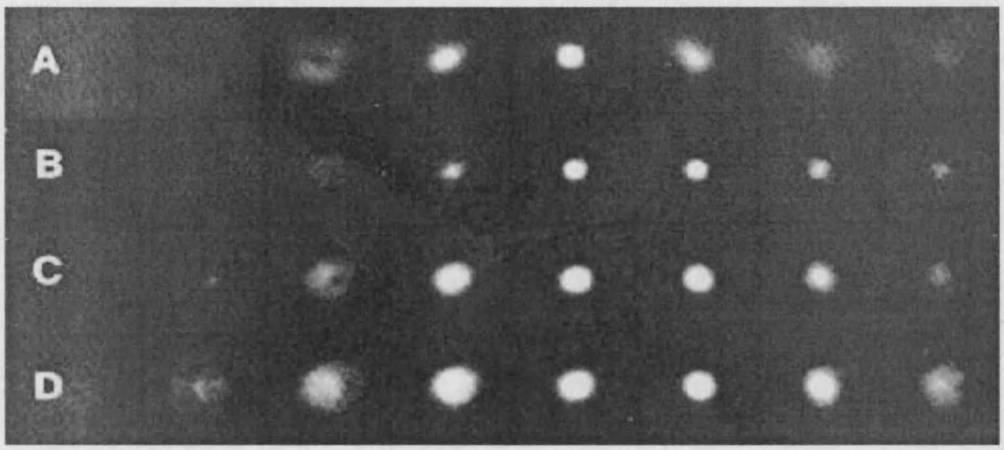

Figure 6-2: The $x - y$ of (a)wide-field PSF. (b) Confocal PSF with minimum detector hole. (c)Confocal PSF with medium detector hole. (d)Confocal PSF with large detector hole in Shaw and Rawlins experiment.

depending on the value of the $z-$scale set. If the $z-$scale is set up to be small then the resulting optical sections are close, and more precise properties of the object are observed. On the other hand if the scale is set to be large then the distance between the layers will be relatively big and the expected loss of information is accordingly high. In our data the vertical scale unit is approximately 4 times the horizontal scale. (A *micron*, μm, is the scale unit used in imaging using confocal microscope and is equal to $10^{-6}m$.)

The scattering of light as it passes through the specimen on the way to the focal point and again on the way back to the detector, gives rise to the point spread function. A point spread function could be thought as the limiting statistical average for the image of many such points in the neighbourhood Λ_i of a single unit i. Shaw and Rawlins (1991) conducted experiment for assessing the point spread function for different types of microscopes. They describe the 2D point spread function as a central disc surrounded by rings of subsidiary maxima as shown in figure $(6 - 2)$). In 3D imaging, the point spread function also has rings which expand away from the focal plane (see figure

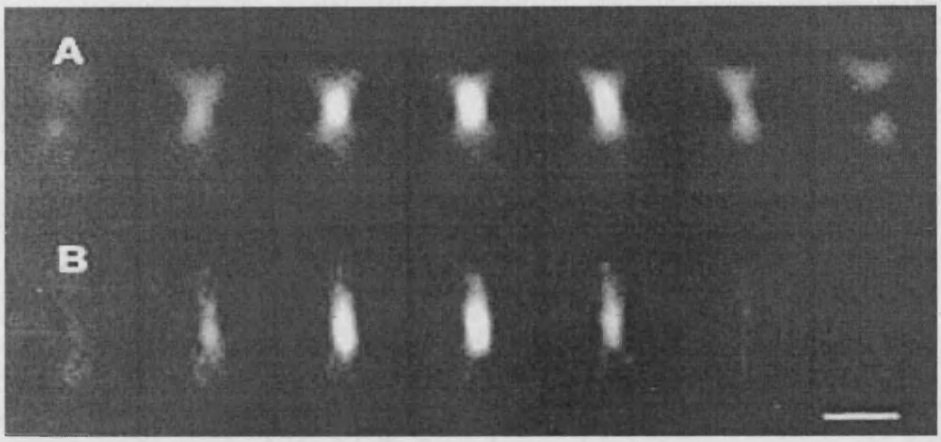

Figure 6-3: $x - z$ sections of PSF of (a) wide-field microscope. (b)Minimum pinhole confocal microscope as in Shaw and Rawlins experiment.

$(6 - 3)$). This description coincides with an ellipsoid shape neighbourhood where the altitude of the ellipsoid is in the direction of the z-axis. Shaw and Rawlins also noticed that the peak width of the PSF ranges from 0.23 to 0.38 μm in the plane and from 0.8 to 1.5 μm in z for confocal microscopes. In our application as the unit scale in the $z-$ direction is set to be larger than the unit scale used for the xy-plane we shall set equal semi-axes for the ellipsoidal shape point spread function. To use the blur effect we shall define Λ_i as the 3D vicinity of a voxel i that contributes to the value of the record for that voxel, Y_i. The shape of the vicinity Λ is approximated by a sphere of diameter 5 voxels length. Each voxel's record in Λ_i will contribute to the value of Y_i with a weight w_i. The sum of the weights of these voxels is set to one. According to the findings of Wilson's experiment (1990) for measuring the point spread function which is shown in figure $(3 - 18)$, high signals are expected from the centre of the vicinity of the focal point and lower responses away from that focal point. Thus a good approximation for the weight kernel is the Gaussian distribution with mean being the centre of the focal voxel and variance v.

Another important aspect to consider when working with records in three dimen-

sional space is the attenuation effect. Optical sections that are close to the surface usually have relatively clear signals, whereas the signals received from the deeper sections are more attenuated (see figure$(6-1)$), this is due to diffraction and absorption of photons as they pass through the sample.

Values of the minimum and maximum record values through consecutive 2D optical sections, (see figure $(6-4)$) indicate that the dark level added to the specimen, d say, is approximately 20. Adding a dark level to the entire specimen being imaged is a common technique used to increase the contrast of the image. This level is not affected by the attenuation factor. Thus to assess the attenuation effect through the optical sections we computed the mean of each 2D section after subtracting the value of $d = 20$. The resulting averages shown in the lower section of figure $(6-4)$ indicate a geometric form of attenuation. That is if a given intensity of dye in the sample produces a signal with mean Ψ at depth k_1, the same intensity at depth k_2 gives a signal with mean $\Psi\Upsilon^{k_2-k_1}$. We estimate the parameter of the geometric function Υ to be 0.945. This estimated geometric effect matches the observed trend very closely (see the lower section of figure $(6-4)$). We shall still assume the records (see section (2.1.2)) $Y_{ijk}, i = 1,\ldots,S_x, j = 1,\ldots,S_y, k = 1,\ldots,S_z$ are conditionally independent given $X = x$, the true 3D image. The mean of this distribution is equal to

$$\mu_{ijk} = \Upsilon^{k-1}(\tilde{\mu}_{ijk} - d) + d, \tag{6.1}$$

where $\tilde{\mu}_{ijk}$ is the mean intensity level of the blurred records which includes the dark background d and is equal to $\sum_{t\in\Lambda_{ijk}} w_t\breve{\mu}_t$. The mean of the un-blurred record $\breve{\mu}_t$ is defined as

$$\breve{\mu}_t = \begin{cases} \tau_0 & \text{if voxel } t \text{ is in the background} \\ \tau_0 + \lambda_l & \text{if voxel } t \text{ is inside the cell } l, \end{cases}$$

where τ_0 is the mean of the background records and λ_l is the intensity level of cell l, assumed constant within that cell. To estimate the variance of the records we compute

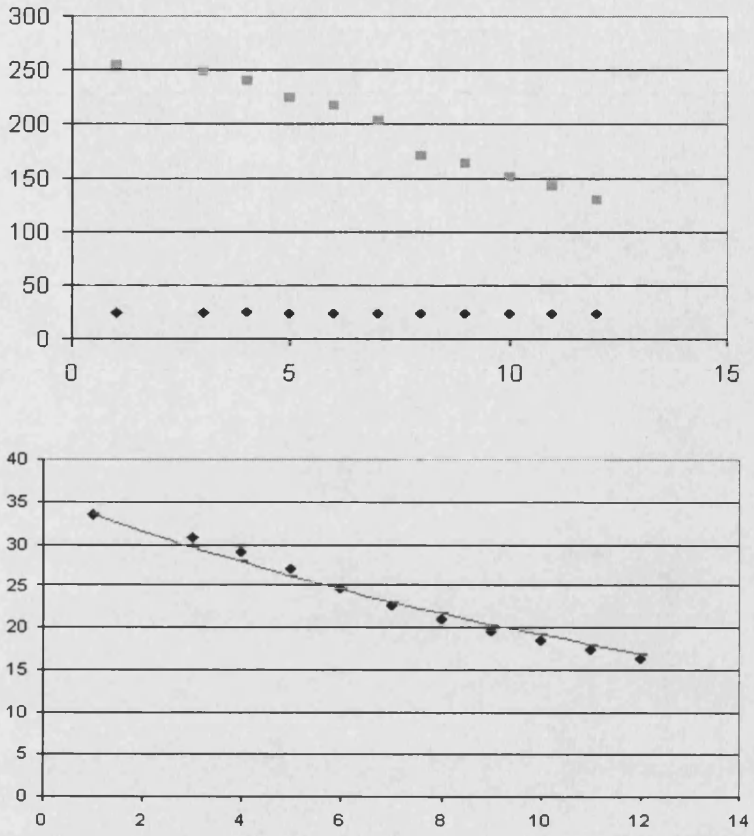

Figure 6-4: The observed maximum(light dots) and minimum (dark dots) record values through consecutive optical sections (top). The average of each optical section after subtracting the dark level (dots) with the estimated geometric fit for these averages (trend) superimposed (bottom).

the samples means and variances for some cells of different optical sections. A plot of these variances against the means is shown in figure (6−5). Consequently the following relation is assumed

$$\sigma_{ijk}^2 = \Upsilon^{2(k-1)}\{\phi_0^2 + c^*(\mu_{ijk} - \tau_0\Upsilon^{k-1})\}, \tag{6.2}$$

where ϕ_0^2 is the variance of the background area and c^* is a positive constant. If the Poisson assumption of the signals holds then the term in braces in equation (6.2)

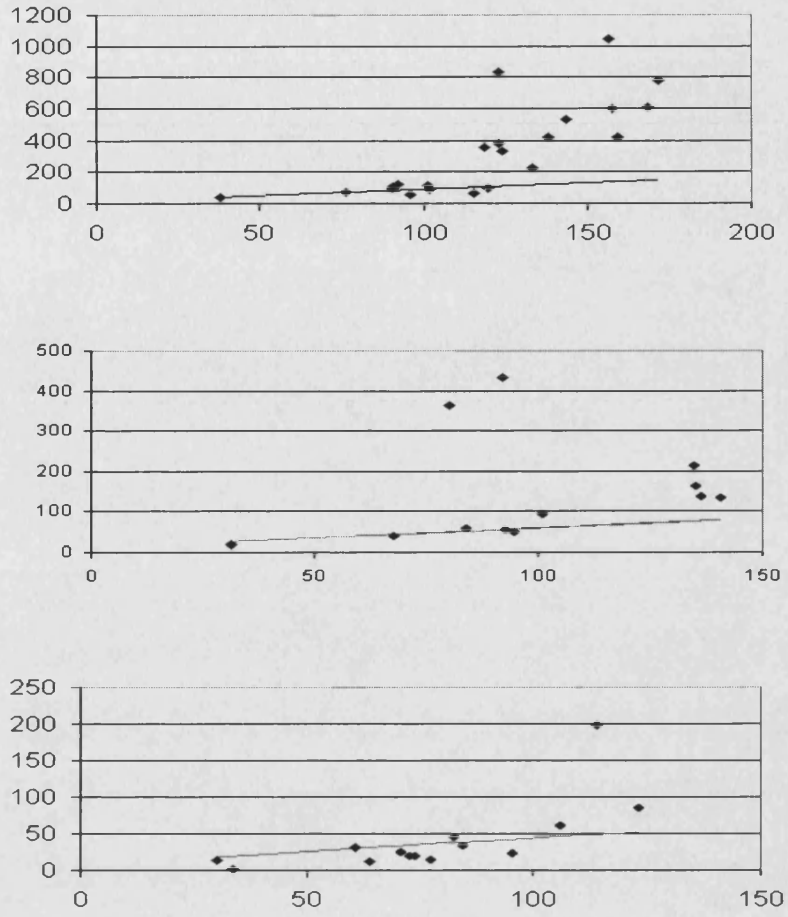

Figure 6-5: Plots of the variances of different cells against their means (dots) for optical sections 3 (top), 7 (middle) and 10 (bottom). The solid lines represent the defined relation between the mean and the variance.

should be multiplied by $\Upsilon^{(k-1)}$ instead of $\Upsilon^{2(k-1)}$ and that will reduce the values of the variance. However, we shall use the observed relation in equation (6.2) obtained from analysis of the data. The values of τ_0 and ϕ_0^2 are estimated using the top 2D section and are set to 42.0 and 46.5 respectively and the value of c^* is set to 0.5 in order to fit closely properties of cells with low mean and low variance. Sample means and variances for cells in different optical sections are shown with the fitted function (6.2) in figure $(6-5)$. The sample sizes are different even for the same cells in different optical sections. The reason is that the cells vary in size, moreover optical sections for

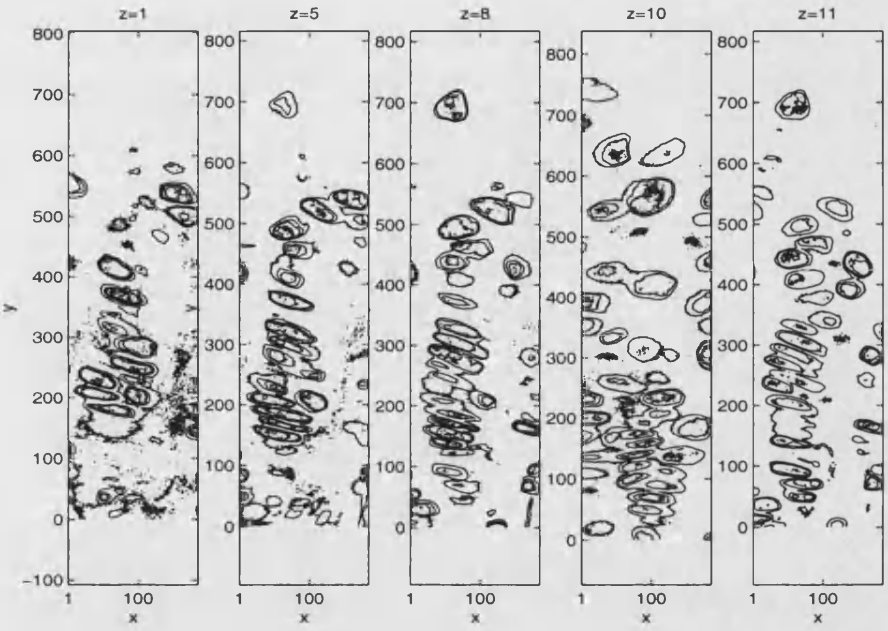

Figure 6-6: Contour plot of different 2D optical sections for the experimental data for optical frames 1, 5, 8, 10 and 11.

the same cells are different because of the nature of cells, the blurring effects, diffraction and attenuation effects of the lights.

It follows that the likelihood function of the attenuated blurred records Y given X is

$$\mathcal{L}(y|x) = \prod_{k=1}^{S_z} \prod_{j=1}^{S_y} \prod_{i=1}^{S_x} \frac{1}{\sqrt{2\pi}\sigma_{ijk}} \exp\{-\frac{(y_{ijk} - \mu_{ijk})^2}{2\sigma_{ijk}^2}\}, \tag{6.3}$$

where each μ_{ijk} and σ_{ijk}^2 is as defined in equations (6.1) and (6.2).

6.3 Defining a 3D object

To describe a typical object in 3D space we viewed data in a number of different 2D planes orthogonal to $x-$axis, $y-$axis or $z-$axis. Figures $(6-6)$, $(6-7)$ and $(6-8)$ are contour plots for these images, with the 2D layers of the specimen arranged neatly and in an increasing order to give a better view of the 3D image. The surfaces of the cells in

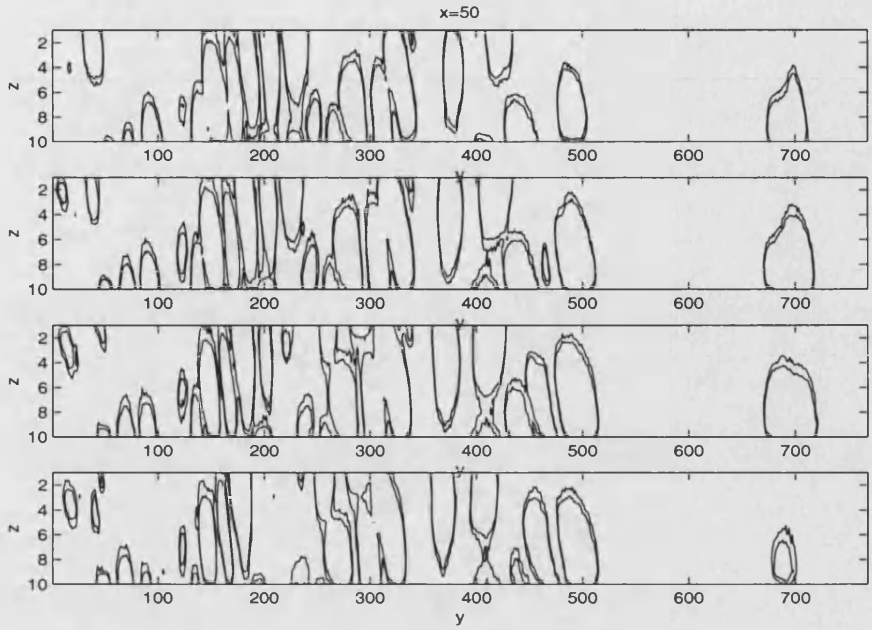

Figure 6-7: Contour plot of 2D sections for the experimental data orthogonal to the $x-$axis. From top to bottom: $Y_{50,j,k}, Y_{60,j,k}, Y_{70,j,k}, Y_{80,j,k}$, $j = 1, \ldots, 768$, $k = 1, \ldots, 10$.

the different optical sections can still be approximated by ellipses. The outlines of the cells shown in figures $(6-6)$ and $(6-7)$ show that the 2D sections are approximately elliptical. Piecing together a series of sections of the same cell gives an ellipsoidal shape and we shall take this as the 3D shape of the cells in our model. The quadratic surface of an ellipsoid is the graph of a second degree equation in x, y and z presented by

$$\frac{(x-c_x)^2}{a^2} + \frac{(y-c_y)^2}{b^2} + \frac{(z-c_z)^2}{c^2} = 1,$$

where (c_x, c_y, c_z) are the coordinates of the centre of the ellipsoids and a, b, c are the semi-axes representing the width, length and height of the ellipsoid.

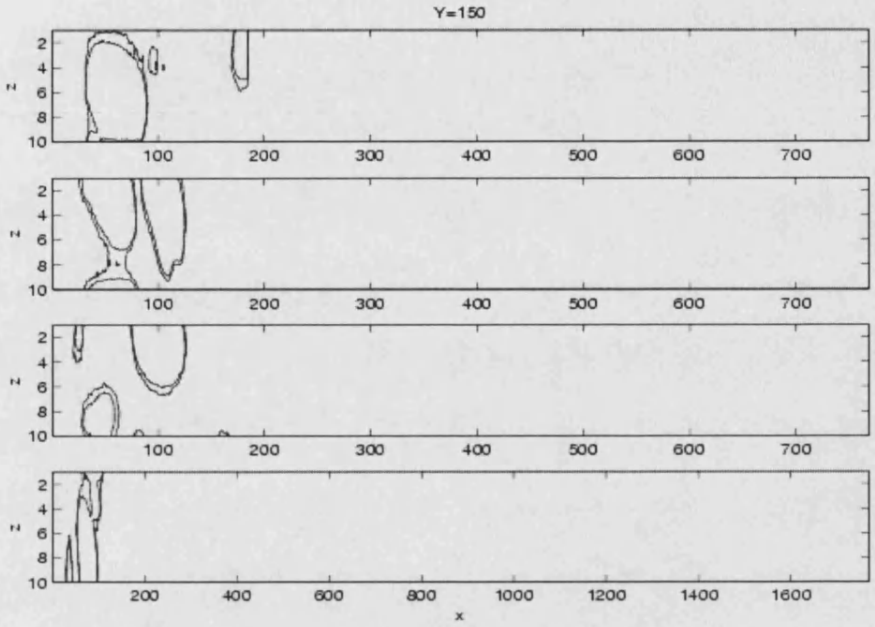

Figure 6-8: Contour plot of 2D sections for the experimental data orthogonal to the $y-$axis. From top to bottom: $Y_{i,150,k}, Y_{i,200,k}, Y_{i,250,k}, Y_{i,300,k}$, $i = 1, \ldots, 184$, $k = 1, \ldots, 10$.

6.4 Orientation of the object

To allow the cells to lie in different orientations in the space we have to consider rotations of the ellipsoids in the 3D space through the *Euler* angles $\theta_1, \theta_2, \theta_3$ (for more details the reader may refer to Spiegel (1967)). Let $P = (p_x, p_y, p_z)$ be the coordinates of a point in the $xyz-$system and let (p'_x, p'_y, p'_z) be the coordinate of P in the new system $x'y'z'$ after rotating through angles $\theta_1, \theta_2, \theta_3$ around the original point $O = (0, 0, 0)$. To calculate (p'_x, p'_y, p'_z) given (p_x, p_y, p_z) the following transformation matrix is used,

$$A = \begin{pmatrix} e_1e_3 - s_1e_2s_3 & s_1e_3 + e_1e_2s_3 & s_2s_3 \\ -e_1s_3 - s_1e_2e_3 & -s_1s_3 + e_1e_2e_3 & s_2e_3 \\ s_1s_2 & -e_1s_2 & e_2 \end{pmatrix}.$$

where $e_i = \cos(\theta_i)$ and $s_i = \sin(\theta_i)$, $i = 1, 2, 3$. We set $(p'_x, p'_y, p'_z)^T = A(p_x, p_y, p_z)^T$. To compute (p_x, p_y, p_z) given (p'_x, p'_y, p'_z) we simply use the transpose of the above matrix, $(p_x, p_y, p_z)^T = A^T(p'_x, p'_y, p'_z)^T$. If the rotation is around a point (c_x, c_y, c_z) other than the origin O we use the transformation of translation of axes first then we apply the transformation of the orientation process, in other words

$$\begin{pmatrix} p'_x \\ p'_y \\ p'_z \end{pmatrix} = A \begin{pmatrix} p_x - c_x \\ p_y - c_y \\ p_z - c_z \end{pmatrix} + \begin{pmatrix} c_x \\ c_y \\ c_z \end{pmatrix}. \tag{6.4}$$

To draw an ellipsoid without rotation we first determine its centre point coordinates (c_x, c_y, c_z) in the 3D $xyz-$space, then set the lengths of its semi-axes, a, b and c. The first semi-axis lies in the direction of the $x-$axis, the second in the direction of the $y-$axis and the third in the direction of the $z-$axis. To rotate the ellipsoid around its centre point we rotate each point covered by the original ellipsoid through the angles of rotation $(\theta_1, \theta_2, \theta_3)$ around (c_x, c_y, c_z) using equation (6.4).

For a point $P = (p_x, p_y, p_z)$ the above rotation system is uniquely determined for $0 \leq \theta_1 < \pi, 0 \leq \theta_2 < 2\pi, 0 \leq \theta_3 < 2\pi$. However, because of the symmetry of the shape of the ellipsoid around its centre point angles $\theta_3 = \alpha$ and $\theta_3 = \alpha + \pi$ take an ellipsoid onto exactly the same set and we get a unique rotation of ellipsoids using angles $0 \leq \theta_1 < \pi, 0 \leq \theta_2 < \pi, 0 \leq \theta_3 < \pi$.

6.5 The object space

Our modified object X_t in the 3D space will be modelled by a pair (l, m) where $l = (c_x, c_y, c_z)$ represents the location of the centre of the ellipsoid and the associated mark is $m = (a, b, c, \theta_1, \theta_2, \theta_3, \lambda)$. The location of the centre l is required to lie within a window L covering the grid of lattice points $\{1, \ldots, S_x\} \times \{1, \ldots, S_y\} \times \{1, \ldots, S_z\}$. We define L as the 3D rectangle window covering the continuous space $[0, L_x) \times [0, L_y) \times [0, L_z)$. Other choices can be used for L as described in section (3.3).

Each semi-axis a, b, c is allowed to take values in the interval $(\min_{axis}, \max_{axis})$. We shall set $\min_{axis} = 3$ and $\max_{axis} = 60$, the same values used in 2D, because we are allowing unconstrained rotations of the objects. The intensity of the cells will lie in the range $M_\lambda = (40, 213)$. The general state space of the mark m is the product of the spaces of its components, $M = M_{axis} \times M_{axis} \times M_{axis} \times (0, \pi) \times (0, \pi) \times (0, \pi) \times M_\lambda$.

The location l takes a value in $L \subset \mathbb{R}^3$ and the mark m takes a value in $M \subset \mathbb{R}^7$, so a single object X_i lies in $\mathbf{U} = L \times M \subset \mathbb{R}^{10}$.

6.6 The prior model

An object configuration is a finite unordered set of objects $X = \{X_1, \ldots, X_N\}$, where $X_i \in \mathbf{U} \subset \mathbb{R}^{10}, i = 1, \ldots, N$. Let $\mathbb{E}$ denote the space of all configurations of X. The formulation of the state X has two components: an index N and a value on the space of the unordered set of N objects, say $\mathcal{U}^N$. The space of $\mathcal{U}^N$ can be defined by considering the space of the ordered set of N objects, $\mathbf{U}^N$ say. The prior distribution of X is defined by reference to the Poisson point process and conditioned on the number of points. The process is defined in detail in section (3.3), noting it is now used for the 3D problem rather than 2D, i.e., the centre points l are from 3D Poisson process with intensity one. We denote by $\mathcal{V}_L$ the volume covered by the 3D window L. This model is then extended to a marked point process where given $N = n$, each point has an independent mark m from ν, a probability measure on M taken such that $\nu(M) = 1$. In $\mathbf{U} = L \times M$

we define a finite measure $\mu = \rho \otimes \nu$ where ρ is Lebesgue measure on L. We refer to this reference measure on $\mathbb{E}$ defined above as $\Gamma^*(X, N)$. With respect to this constructive reference measure $\Gamma^*(X, N)$ we define the distribution of the object configurations in $\mathbb{E}$. We shall continue using the hard core model as the prior model for (X, N) which has a density,

$$\pi(x, n) = k\beta^{n(x)} \prod_{\text{all pairs } (x_i, x_j)} I[\text{no overlap between the objects } x_i \text{ and } x_j], \quad (6.5)$$

with respect to $\Gamma^*(X, N)$. Here k is a normalising constant which would be computed by integrating over all the values of X and N in $\mathbb{E}$, and β is a constant reflecting the intensity of the cell process.

The space of an object $\mathbf{U}$ is a subset of $\mathbb{R}^{10}$, so $\mu(x_i)$ is a Lebesgue measure. Therefore, in the reference measure, the vector $(c_{xi}, c_{yi}, c_{zi}, a_i, b_i, c_i, \theta_{1i}, \theta_{2i}, \theta_{3i}, \lambda_i)$ of a single object x_i has a density, $f_{X_i}(x_i)$, with respect to Lebesgue measure on $\mathbb{R}^{10}$. In the part of $\mathbb{E}$ corresponding to $N = n$ objects in the scene, the sub-reference measure $\Gamma_n^*(X)$ can be written as a sub-density of the n objects and the unordered set $\{x_1, x_2, \ldots, x_n\}$ with respect to the Lebesgue measure on $\mathbb{R}^{10n}$ and is equal to

$$e^{-\mathcal{V}_L} \mathcal{V}_L^n \prod_{i=1}^{n} f_{X_i}(x_i). \quad (6.6)$$

This sub-density is to be used in the evaluation of the acceptance probabilities for the MCMC algorithm as we shall see in the next section.

To model the variability of the marks we define suitable distributions with respect to ν on M. The semi-axes a, b and c are assumed to be independent, each defined on M_{axis} and their density is proportional to that of multivariate normal with mean

(μ_a, μ_b, μ_c) and covariance matrix

$$\Sigma = \begin{bmatrix} \gamma^2 & 0 & 0 \\ 0 & \gamma^2 & 0 \\ 0 & 0 & \gamma_c^2 \end{bmatrix}.$$

The angles of rotation $\theta_i, i = 1,2,3$ are assumed to be independent, each having a probability density

$$\pi_{\theta_i}(\theta_i) = \frac{|\cos(\theta_i)| + \frac{1}{\pi}}{3}, \quad 0 \leq \theta_i < \pi.$$

Finally the intensity rate λ is assumed to be uniformly distributed over $(\min_\lambda, \max_\lambda)$. So for $\nu(m)$, the mark distribution has density with respect to Lebesgue measure on $\mathbb{R}^7$

$$z\frac{|\cos(\theta_1)| + \frac{1}{\pi}}{3}\frac{|\cos(\theta_2)| + \frac{1}{\pi}}{3}\frac{|\cos(\theta_3)| + \frac{1}{\pi}}{3}\exp(-\frac{1}{2\gamma^2}((a-\mu_a)^2 + (b-\mu_b)^2) - \frac{1}{2\gamma_c^2}(c-\mu_c)^2)\frac{1}{\max_\lambda - \min_\lambda}I_{[a,b,c \in M_{axis}]}, \tag{6.7}$$

where z is a normalising constant computed by integrating over all $m \in M$ with respect to the Lebesgue measure ν on M. For a single object $x_i = (l_i, m_i)$ in the scene we can define the density $f_{X_i}(x_i)$ on $\mathbb{R}^{10}$ to be

$$f_{X_i}(c_x, c_y, c_z, a, b, c, \theta_1, \theta_2, \theta_3, \lambda) = \frac{z}{\mathcal{V}_L}\frac{|\cos(\theta_1)| + \frac{1}{\pi}}{3}\frac{|\cos(\theta_2)| + \frac{1}{\pi}}{3}\frac{|\cos(\theta_3)| + \frac{1}{\pi}}{3}$$
$$\exp(-\frac{1}{2\gamma^2}((a-\mu_a)^2 + (b-\mu_b)^2) - \frac{1}{2\gamma_c^2}(c-\mu_c)^2)\frac{1}{\max_\lambda - \min_\lambda}I_{[a,b,c \in M_{axis}]}. \tag{6.8}$$

Substituting for f_{X_i} in equation (6.6) we get that for $N = n$ objects in $\mathbb{E}$, the part of the prior distribution of an object configuration with respect to Lebesgue measure is proportional to

$$\frac{\beta^n}{(\max_\lambda - \min_\lambda)^n}\prod_{t=1}^{n}\{\frac{|\cos(\theta_{1,t})| + \frac{1}{\pi}}{3}\frac{|\cos(\theta_{2,t})| + \frac{1}{\pi}}{3}\frac{|\cos(\theta_{3,t})| + \frac{1}{\pi}}{3}$$

$$\exp(-\frac{1}{2\gamma^2}((a_t-\mu_a)^2+(b_t-\mu_b)^2)-\frac{1}{2\gamma_c^2}(c_t-\mu_c)^2)I_{[a_t,b_t,c_t\in M_{axis}]}\}$$
$$I[\text{no overlap between the cells in the scene}], \qquad (6.9)$$

6.7 MCMC algorithm

To implement the MCMC simulation we consider birth and death moves, merge and split moves, shifts, resizes, changes in orientation and finally updating the intensity level. We use the random scheme update where at each iteration we choose a move of type $v = \{1, 2, \ldots, 8\}$ with probability $j_v(x)$ where $v = 1$ stands for birth, 2 for death, 3 for shift, 4 for resize, 5 for rotation, 6 for updating the intensity, 7 for a merge move and 8 for a split move. The probability of proposing a move depends on the state of current configuration x; for example if $x = \emptyset$ the only available move is the birth move, i.e., $j_1(x) = 1$.

In analogy to the 2D implementations of the birth, death, shift, resize, rotation, and updating the intensity moves we generalise the proposal kernels and the likelihood function. Accordingly we modify the acceptance rate for these moves with respect to the 3D space. More details about implementing some of these moves will be given in the next chapter. However, in the next section we shall give more detailed information about implementing the 3D split and merge moves.

6.7.1 Split and merge moves

Merging

To attempt to merge two randomly selected cells from the current configuration, say x_A and x_B, the cells have to satisfy several criteria: The Euclidian distance between their centres,

$$\sqrt{(c_{xA}-c_{xB})^2+(c_{yA}-c_{yB})^2+(c_{zA}-c_{zB})^2}$$

should be less than a fixed value ϵ_d; the difference of their intensity should not exceed some fixed value ϵ_λ chosen to be $(\max_\lambda - \min_\lambda)/5$; moreover, the differences between their relative angles of rotation θ_1, θ_2 and θ_3 should not exceed $\epsilon_{\theta_1}, \epsilon_{\theta_2}$ and ϵ_{θ_3} respectively. That is

$$|\theta_{j,A} - \theta_{j,B}| \leq \epsilon_{\theta_j} \text{ or } |\theta_{j,A} - \theta_{j,B}| \geq \pi - \epsilon_{\theta_j}, \quad j = 1,2,3,$$

where $\epsilon_{\theta_j} = \pi/40$, $j = 1,2,3$. If these conditions are satisfied then we attempt to replace x_A and x_B by a new cell x_C. The mechanism of the merge move can be described as follows.

We allocate the centre of the new cell, $l_C = (c_{xC}, c_{yC}, c_{zC})$ on the centre of l_A and l_B,

$$c_{xC} = \frac{c_{xA}+c_{xB}}{2} \qquad c_{yC} = \frac{c_{yA}+c_{yB}}{2} \qquad c_{zC} = \frac{c_{zA}+c_{zB}}{2}. \tag{6.10}$$

In order for the proposed merge to have a high degree of acceptance, the proposed cell C should cover most of the space covered by x_A and x_B individually. To do that we shall propose the semi-axes conditioned on the locations l_A, l_B and on the semi-axes of x_A, x_B using the *spherical polar coordinate system.* Before we proceed we shall give a brief description of this coordinate system.

Let (p_x, p_y, p_z) be the Cartesian coordinates of a point P in the 3D space with respect to the origin O. Let d be the distance from P to O, and Δ_1 be the polar angle associated with the projection P' of P on the $xy-$plane, i.e., Δ_1 is the angle between the positive $x-$axis and OP'. Finally define Δ_2 to be the angle from the positive $z-$axis to OP. Then (d, Δ_1, Δ_2) are the spherical coordinates of P and the graph of the equation $d = c$ where $c > 0$ is a sphere with centre at O. For $P = (p_x, p_y, p_z)$ in Cartesian coordinates,

$$d = \sqrt{p_x^2 + p_y^2 + p_z^2},$$

$$\Delta_1 = \arctan \frac{p_y}{p_x},$$

$$\Delta_2 = \arccos \frac{p_z}{d}.$$

To compute the spherical coordinates of location l_B relative to l_A, we translate l_A to O and let this determine the resulting coordinates of the second point, i.e., we define $P = l_A - l_B = (c_{xA} - c_{xB}, c_{yA} - c_{yB}, c_{zA} - c_{zB})$. Thus

$$d = \sqrt{(c_{xA} - c_{xB})^2 + (c_{yA} - c_{yB})^2 + (c_{zA} - c_{zB})^2},$$

$$\Delta_1 = \arctan \frac{c_{yA} - c_{yB}}{c_{xA} - c_{xB}},$$

$$\Delta_2 = \arccos \frac{c_{zA} - c_{zB}}{d}.$$

There are different positions that the two cells x_A and x_B can take when they are proposed to merge. The choice of a_C, b_C and c_C should be conditional on these positions. For instance if x_A and x_B occupy the same $xy-$region but different $z-$regions then a good region for x_C to cover is the same region of the $xy-$plane but with a height covering both the heights of x_A and x_B. On the other hand if x_A and x_B cover the same vertical space but different regions in the $xy-$plane then the previous region of x_C is not suitable. A better proposal is a cell occupying the same vertical space as x_A and an appropriate horizontal space to replace that of x_A and x_B. These two cases illustrate our objectives in choosing the location l_C and semi-axis lengths a_C, b_C and c_C. The technical implementation for choosing these semi-axes is associated with the polar coordinates which defined earlier. We set,

$$\begin{aligned} a_C &= \tfrac{1}{2}(1 + |\cos(\Delta_1)|)|\sin(\Delta_2)|(a_A + a_B), \\ b_C &= \tfrac{1}{2}(1 + |\sin(\Delta_1)|)|\sin(\Delta_2)|(b_A + b_B), \\ c_C &= \tfrac{1}{2}(1 + |\cos(\Delta_2)|)(c_A + c_B), \end{aligned} \tag{6.11}$$

To ensure that $a_C, b_C, c_C \in (\min_{axis}, \max_{axis})$, we require that

$$
\begin{aligned}
a_A + a_B &\in \left(\frac{2\min_{axis}}{|\sin(\Delta_2)|(|\cos(\Delta_1)|+1)}, \frac{2\max_{axis}}{|\sin(\Delta_2)|(|\cos(\Delta_1)|+1)} \right) \\
b_A + b_B &\in \left(\frac{2\min_{axis}}{|\sin(\Delta_2)|(|\sin(\Delta_1)|+1)}, \frac{2\max_{axis}}{|\sin(\Delta_2)|(|\sin(\Delta_1)|+1)} \right) \\
c_A + c_B &\in \left(\frac{2\min_{axis}}{|\cos(\Delta_2)|+1}, \frac{2\max_{axis}}{|\cos(\Delta_2)|+1} \right).
\end{aligned} \tag{6.12}
$$

The previous conditions prohibit merging of large cells with semi-axis lengths close to the upper limit allowable. The combining operation must be done such that the reverse split move can be constructed to satisfy the detailed balance condition. Hence, if the conditions under which a split move is allowable are not met then the proposed merge move is rejected for the current x_A and x_B cells.

To find the angles of rotation $\theta_{j,C}, j = 1,2,3$ and the intensity λ_C, this part of the merge move will be deterministic. We set $\theta_{j,C}$ to be

$$
\theta_{j,C} = \begin{cases} \frac{1}{2}(\theta_{j,A} + \theta_{j,B}) & \text{if } \epsilon_{\theta_j} > |\theta_{j,A} - \theta_{j,B}| \\ (\frac{1}{2}(\pi + \theta_{j,A} + \theta_{j,B})) \bmod \pi & \text{if } \pi - \epsilon_{\theta_j} < |\theta_{j,A} - \theta_{j,B}|, \end{cases}
$$

for $j = 1,2,3$. The intensity level of the new object is set to be the average value of λ_A and λ_B.

Splitting

For the reverse split move we need to define two points l_A, l_B satisfying the equation $l_C = (l_A + l_B)/2$. This is implemented by drawing one centre say l_A uniformly in a sphere with diameter ϵ_d, and letting this determine the centre of the other object l_B by fixing the midpoint of l_A and l_B to be l_C. That is we draw Δ_1 and Δ_2 uniformly over $(0, 2\pi)$, and also r uniformly over $(0, \frac{\epsilon_d}{2})$, then set

$$
\begin{aligned}
c_{xA} &= c_{xC} + r\cos(\Delta_1)\sin(\Delta_2) & c_{xB} &= c_{xC} - r\cos(\Delta_1)\sin(\Delta_2) \\
c_{yA} &= c_{yC} + r\sin(\Delta_1)\sin(\Delta_2) & c_{yB} &= c_{yC} - r\sin(\Delta_1)\sin(\Delta_2) \\
c_{zA} &= c_{zC} + r\cos(\Delta_2) & c_{zA} &= c_{zC} - r\cos(\Delta_2)
\end{aligned} \tag{6.13}
$$

Δ_2	$0 < \Delta_1 < \frac{\pi}{2}$	$\frac{\pi}{2} < \Delta_1 < \pi$	$\pi < \Delta_1 < \frac{3\pi}{2}$	$\frac{3\pi}{2} < \Delta_1 < 2\pi$
$0 < \Delta_2 < \frac{\pi}{2}$	$\min\{\frac{\epsilon_d}{2}, \frac{c_{zC}}{\cos(\Delta_2)}, \frac{L_x - c_{xC}}{\cos(\Delta_1)\sin(\Delta_2)}, \frac{L_y - c_{yC}}{\sin(\Delta_1)\sin(\Delta_2)}, \frac{c_{yC}}{\sin(\Delta_1)\sin(\Delta_2)}, \frac{L_z - c_{zC}}{\cos(\Delta_2)}, \frac{c_{xC}}{\cos(\Delta_1)\sin(\Delta_2)}\}$	$\min\{\frac{\epsilon_d}{2}, \frac{c_{zC}}{\cos(\Delta_2)}, \frac{-L_x + c_{xC}}{\cos(\Delta_1)\sin(\Delta_2)}, \frac{L_y - c_{yC}}{\sin(\Delta_1)\sin(\Delta_2)}, \frac{c_{yC}}{\sin(\Delta_1)\sin(\Delta_2)}, \frac{L_z - c_{zC}}{\cos(\Delta_2)}, \frac{-c_{xC}}{\cos(\Delta_1)\sin(\Delta_2)}\}$	$\min\{\frac{\epsilon_d}{2}, \frac{c_{zC}}{\cos(\Delta_2)}, \frac{-L_x + c_{xC}}{\cos(\Delta_1)\sin(\Delta_2)}, \frac{-L_y + c_{yC}}{\sin(\Delta_1)\sin(\Delta_2)}, \frac{-c_{yC}}{\sin(\Delta_1)\sin(\Delta_2)}, \frac{L_z - c_{zC}}{\cos(\Delta_2)}, \frac{-c_{xC}}{\cos(\Delta_1)\sin(\Delta_2)}\}$	$\min\{\frac{\epsilon_d}{2}, \frac{c_{zC}}{\cos(\Delta_2)}, \frac{L_x - c_{xC}}{\cos(\Delta_1)\sin(\Delta_2)}, \frac{-L_y + c_{yC}}{\sin(\Delta_1)\sin(\Delta_2)}, \frac{-c_{yC}}{\sin(\Delta_1)\sin(\Delta_2)}, \frac{L_z - c_{zC}}{\cos(\Delta_2)}, \frac{c_{xC}}{\cos(\Delta_1)\sin(\Delta_2)}\}$
$\frac{\pi}{2} < \Delta_2 < \pi$	$\min\{\frac{\epsilon_d}{2}, \frac{-c_{zC}}{\cos(\Delta_2)}, \frac{L_x - c_{xC}}{\cos(\Delta_1)\sin(\Delta_2)}, \frac{L_y - c_{yC}}{\sin(\Delta_1)\sin(\Delta_2)}, \frac{c_{yC}}{\sin(\Delta_1)\sin(\Delta_2)}, \frac{-L_z + c_{zC}}{\cos(\Delta_2)}, \frac{c_{xC}}{\cos(\Delta_1)\sin(\Delta_2)}\}$	$\min\{\frac{\epsilon_d}{2}, \frac{-c_{zC}}{\cos(\Delta_2)}, \frac{-L_x + c_{xC}}{\cos(\Delta_1)\sin(\Delta_2)}, \frac{L_y - c_{yC}}{\sin(\Delta_1)\sin(\Delta_2)}, \frac{c_{yC}}{\sin(\Delta_1)\sin(\Delta_2)}, \frac{-L_z + c_{zC}}{\cos(\Delta_2)}, \frac{-c_{xC}}{\cos(\Delta_1)\sin(\Delta_2)}\}$	$\min\{\frac{\epsilon_d}{2}, \frac{-c_{zC}}{\cos(\Delta_2)}, \frac{-L_x + c_{xC}}{\cos(\Delta_1)\sin(\Delta_2)}, \frac{-L_y + c_{yC}}{\sin(\Delta_1)\sin(\Delta_2)}, \frac{-c_{yC}}{\sin(\Delta_1)\sin(\Delta_2)}, \frac{-L_z + c_{zC}}{\cos(\Delta_2)}, \frac{-c_{xC}}{\cos(\Delta_1)\sin(\Delta_2)}\}$	$\min\{\frac{\epsilon_d}{2}, \frac{-c_{zC}}{\cos(\Delta_2)}, \frac{L_x - c_{xC}}{\cos(\Delta_1)\sin(\Delta_2)}, \frac{-L_y + c_{yC}}{\sin(\Delta_1)\sin(\Delta_2)}, \frac{-c_{yC}}{\sin(\Delta_1)\sin(\Delta_2)}, \frac{-L_z + c_{zC}}{\cos(\Delta_2)}, \frac{c_{xC}}{\cos(\Delta_1)\sin(\Delta_2)}\}$
$\pi < \Delta_2 < \frac{3\pi}{2}$	$\min\{\frac{\epsilon_d}{2}, \frac{-c_{zC}}{\cos(\Delta_2)}, \frac{-L_x + c_{xC}}{\cos(\Delta_1)\sin(\Delta_2)}, \frac{-L_y + c_{yC}}{\sin(\Delta_1)\sin(\Delta_2)}, \frac{-c_{yC}}{\sin(\Delta_1)\sin(\Delta_2)}, \frac{-L_z + c_{zC}}{\cos(\Delta_2)}, \frac{c_{xC}}{\cos(\Delta_1)\sin(\Delta_2)}\}$	$\min\{\frac{\epsilon_d}{2}, \frac{-c_{zC}}{\cos(\Delta_2)}, \frac{L_x - c_{xC}}{\cos(\Delta_1)\sin(\Delta_2)}, \frac{-L_y + c_{yC}}{\sin(\Delta_1)\sin(\Delta_2)}, \frac{-c_{yC}}{\sin(\Delta_1)\sin(\Delta_2)}, \frac{-L_z + c_{zC}}{\cos(\Delta_2)}, \frac{c_{xC}}{\cos(\Delta_1)\sin(\Delta_2)}\}$	$\min\{\frac{\epsilon_d}{2}, \frac{-c_{zC}}{\cos(\Delta_2)}, \frac{L_x - c_{xC}}{\cos(\Delta_1)\sin(\Delta_2)}, \frac{L_y - c_{yC}}{\sin(\Delta_1)\sin(\Delta_2)}, \frac{c_{yC}}{\sin(\Delta_1)\sin(\Delta_2)}, \frac{-L_z + c_{zC}}{\cos(\Delta_2)}, \frac{c_{xC}}{\cos(\Delta_1)\sin(\Delta_2)}\}$	$\min\{\frac{\epsilon_d}{2}, \frac{-c_{zC}}{\cos(\Delta_2)}, \frac{-L_x + c_{xC}}{\cos(\Delta_1)\sin(\Delta_2)}, \frac{L_y - c_{yC}}{\sin(\Delta_1)\sin(\Delta_2)}, \frac{c_{yC}}{\sin(\Delta_1)\sin(\Delta_2)}, \frac{-L_z + c_{zC}}{\cos(\Delta_2)}, \frac{-c_{xC}}{\cos(\Delta_1)\sin(\Delta_2)}\}$
$0 < \Delta_2 < \frac{\pi}{2}$	$\min\{\frac{\epsilon_d}{2}, \frac{c_{zC}}{\cos(\Delta_2)}, \frac{-L_x + c_{xC}}{\cos(\Delta_1)\sin(\Delta_2)}, \frac{-L_y + c_{yC}}{\sin(\Delta_1)\sin(\Delta_2)}, \frac{-c_{yC}}{\sin(\Delta_1)\sin(\Delta_2)}, \frac{L_z - c_{zC}}{\cos(\Delta_2)}, \frac{-c_{xC}}{\cos(\Delta_1)\sin(\Delta_2)}\}$	$\min\{\frac{\epsilon_d}{2}, \frac{c_{zC}}{\cos(\Delta_2)}, \frac{L_x - c_{xC}}{\cos(\Delta_1)\sin(\Delta_2)}, \frac{-L_y + c_{yC}}{\sin(\Delta_1)\sin(\Delta_2)}, \frac{-c_{yC}}{\sin(\Delta_1)\sin(\Delta_2)}, \frac{L_z - c_{zC}}{\cos(\Delta_2)}, \frac{c_{xC}}{\cos(\Delta_1)\sin(\Delta_2)}\}$	$\min\{\frac{\epsilon_d}{2}, \frac{c_{zC}}{\cos(\Delta_2)}, \frac{L_x - c_{xC}}{\cos(\Delta_1)\sin(\Delta_2)}, \frac{L_y - c_{yC}}{\sin(\Delta_1)\sin(\Delta_2)}, \frac{c_{yC}}{\sin(\Delta_1)\sin(\Delta_2)}, \frac{L_z - c_{zC}}{\cos(\Delta_2)}, \frac{c_{xC}}{\cos(\Delta_1)\sin(\Delta_2)}\}$	$\min\{\frac{\epsilon_d}{2}, \frac{c_{zC}}{\cos(\Delta_2)}, \frac{-L_x + c_{xC}}{\cos(\Delta_1)\sin(\Delta_2)}, \frac{L_y - c_{yC}}{\sin(\Delta_1)\sin(\Delta_2)}, \frac{c_{yC}}{\sin(\Delta_1)\sin(\Delta_2)}, \frac{L_z - c_{zC}}{\cos(\Delta_2)}, \frac{-c_{xC}}{\cos(\Delta_1)\sin(\Delta_2)}\}$

Table 6.1: The upper bounds of r conditioned on Δ_1, Δ_2.

where $(c_{xi}, c_{y_i}, c_{zi})$ are the Cartesian coordinates of l_i and have to lie within the window L. Although in principle we require r to be less than or equal to $\epsilon_d/2$, the condition $l_i \in L$ will put more constraints on the range of r. The choice of the radius r will be conditioned on Δ_1 and Δ_2. The lower bound $\min_r$ of the radius r is $\min_{axis}$, while the upper bound $\max_r$ of r can be computed using table (6.1). Unfortunately the uniform choice of r, Δ_1 and Δ_2 slows down the rate of proposing good candidates for l_A and l_B. In the next chapter we shall describe modifications for the proposal densities of these variables to overcome this speed problem.

There should be a bijection between the merged cell and the two split cells. That is the process for determining the semi-axes for the split cells should be done in such a way that it is possible for the resulting cells to satisfy the merging conditions. Therefore to determine the semi-axes for the new cells we generate three random variables u_a, u_b and u_c from normal distributions with mean zero and variance ϖ^2, and set

$$\begin{array}{rclrcl} a_A &=& \frac{a_C}{|\sin(\Delta_2)|(|\cos(\Delta_1)|+1)} + u_a, & a_B &=& \frac{a_C}{|\sin(\Delta_2)|(|\cos(\Delta_1)|+1)} - u_a \\ b_A &=& \frac{b_C}{|\sin(\Delta_2)|(|\sin(\Delta_1)|+1)} + u_b, & b_B &=& \frac{b_C}{|\sin(\Delta_2)|(|\sin(\Delta_1)|+1)} - u_b \\ c_A &=& \frac{c_C}{|\cos(\Delta_2)|+1} + u_c, & c_B &=& \frac{c_C}{|\cos(\Delta_2)|+1} - u_c. \end{array} \tag{6.14}$$

The choice of the parameters of the distributions of u_a, u_b and u_c controls the difference between the relative semi-axes of the two split cells. By choosing their mean to be zero and small variances we encourage the two cells to have approximately equal corresponding axes. To ensure the semi-axes lie within $M_{axis} = [3, 60]$ the variables u_a, u_b and u_c have to lie in the following ranges

$$u_a \in (\min_{u_a}, \max_{u_a}),$$

$$u_b \in (\min_{u_b}, \max_{u_b}),$$

$$u_c \in (\min_{u_c}, \max_{u_c}),$$

where

$$\max_{u_a} = \max\{3 - \frac{a_C}{|\sin(\Delta_2)|(|\cos(\Delta_1)|+1)}, -60 + \frac{a_C}{|\sin(\Delta_2)|(|\cos(\Delta_1)|+1)}\},$$

$$\max_{u_b} = \max\{3 - \frac{b_C}{|\sin(\Delta_2)|(|\sin(\Delta_1)|+1)}, -60 + \frac{b_C}{|\sin(\Delta_2)|(|\sin(\Delta_1)|+1)}\},$$

$\max_{u_c} = \max\{3 - \frac{c_C}{|\cos(\Delta_2)|+1}, -60 + \frac{c_C}{|\cos(\Delta_2)|+1}\}$, and $\min_{u_i} = -\max_{u_i}$, $i = a, b, c$.

For the angles of rotation we draw three random variables u_{θ_i}, $i = 1, 2, 3$ uniformly

from $(-\frac{\epsilon_{\theta_i}}{2}, \frac{\epsilon_{\theta_i}}{2})$ and set

$$\theta_{i,A} = \theta_{i,C} - u_{\theta_i} \qquad \theta_{i,B} = \theta_{i,C} + u_{\theta_i} \;\; i = 1, 2, 3. \tag{6.15}$$

The same procedure is applied for determining the intensity level of cell x_A and x_B, where we define a random variable u_λ uniformly over $(\min_{u_\lambda}, \max_{u_\lambda})$ and set

$$\lambda_A = \lambda_C - u_\lambda \qquad \lambda_B = \lambda_C + u_\lambda, \tag{6.16}$$

where

$$\max\{-\frac{\epsilon_\lambda}{2}, \lambda_C - \max_\lambda, \min_\lambda - \lambda_C\} \leq u_\lambda \leq \max_{u_\lambda} = \min\{\frac{\epsilon_\lambda}{2}, \lambda_C - \min_\lambda, \max_\lambda - \lambda_C\}.$$

To calculate the acceptance probability associated with the merge and split moves, suppose that the current state is $x = \{x_1, \ldots, x_{n-1}, x_n, x_{n+1}\}$ and we propose to combine two randomly selected cells say x_n, x_{n+1} to x_n^* which if accepted will take the current state to $x' = \{x_1, \ldots, x_{n-1}, x_n^*\}$. The acceptance probability of the merge move labelled $v = 7$ which that takes the state from $x = \{x_1, \ldots, x_{n-1}, x_n, x_{n+1}\}$ to $x' = \{x_1, \ldots, x_{n-1}, x_n^*\}$ becomes

$$\min\left\{1, \frac{\pi(x', n)q(u)(n+2)j_8(x')}{\pi(x, n+1)2j_7(x)} \left|\frac{d(u, x}{d(x')}\right| \frac{\mathcal{L}(y|x', n)}{\mathcal{L}(y|x, n+1)}\right\}, \tag{6.17}$$

where u is a vector of random variables $u = (u_\lambda, u_{\theta_1}, u_{\theta_2}, u_{\theta_3}, \Delta_1, \Delta_2, r, u_a, u_b, u_c)$. The above acceptance probability can be written as the product of four terms.

1. The first term is

$$\frac{1}{\beta} \frac{\pi_{a,b,c}(a_n^*, b_n^*, c_n^*)}{\pi_{a,b,c}(a_n, b_n, c_n)\pi_{a,b,c}(a_{n+1}, b_{n+1}, c_{n+1})} \frac{\pi_{\theta_1}(\theta_{1,n^*})}{\pi_{\theta_1}(\theta_{1,n})\pi_{\theta_1}(\theta_{1,n+1})}$$

$$\frac{\pi_{\theta_2}(\theta_{2,n^*})}{\pi_{\theta_2}(\theta_{2,n})\pi_{\theta_2}(\theta_{2,n+1})}\frac{\pi_{\theta_3}(\theta_{3,n^*})}{\pi_{\theta_3}(\theta_{3,n})\pi_{\theta_3}(\theta_{3,n+1})}\frac{\pi_{\lambda}(\lambda_n^*)}{\pi_{\lambda}(\lambda_n)\pi_{\lambda}(\lambda_{n+1})}I[\text{no overlap}],$$

which is equal to

$$\frac{z_1^{-1}(3\pi)^3(2\pi)^{\frac{3}{2}}\gamma^2\gamma_c(\pi|\cos(\theta_{1,n^*})|+1)(\pi|\cos(\theta_{2,n^*})|+1)}{\beta(\pi|\cos(\theta_{1,n})|+1)(\pi|\cos(\theta_{1,n+1})|+1)(\pi|\cos(\theta_{2,n})|+1)}$$

$$\frac{(\pi|\cos(\theta_{3,n^*})|+1)(\max_\lambda-\min_\lambda)}{(\pi|\cos(\theta_{2,n+1})|+1)(\pi|\cos(\theta_{3,n})|+1)(\pi|\cos(\theta_{3,n+1})|+1)}\exp(-$$

$$\frac{1}{2\gamma^2}[(a_{n^*}-\mu_a)^2+(b_{n^*}-\mu_b)^2-(a_n-\mu_a)^2-(b_n-\mu_b)^2-(a_{n+1}$$
$$-\mu_a)^2-(b_{n+1}-\mu_b)^2]-\frac{1}{2\gamma_c^2}[(c_{n^*}-\mu_c)^2-(c_n-\mu_c)^2-(c_{n+1}-$$
$$\mu_c)^2])I[\text{no overlap}]I[\min_{axis}\leq a_i,b_i,c_i,\ i=n^*,n,n+1\leq\max_{axis}], \qquad (6.18)$$

where z_1 is the normalising constant for the density of the semi-axes and is equal to

$$\int_{min_{axis}}^{max_{axis}}\int_{min_{axis}}^{max_{axis}}\int_{min_{axis}}^{max_{axis}}\frac{1}{(\sqrt{2\pi})^3\gamma^2\gamma_c}\exp(-\frac{1}{2\gamma^2}[(w-\mu_a)^2+(y-\mu_b)^2]$$
$$-\frac{1}{2\gamma_c^2}(z-\mu_c)^2)dwdydz]^{-1}.$$

2. The second term is

$$\frac{n+1}{2}\frac{j_{split}(x')}{j_{merge}(x)}q_{u_\lambda}(u_\lambda)q_{u_{\theta_1}}(u_{\theta_1})q_{u_{\theta_2}}(u_{\theta_2})q_{u_{\theta_3}}(u_{\theta_3})q_{\Delta_1}(\Delta_1)$$
$$q_{\Delta_2}(\Delta_2)q_{r|\Delta_1,\Delta_2}(r)q_{u_a}(u_a)q_{u_b}(u_b)q_{u_c}(u_c).$$

Substituting for the specified distributions, the term becomes

$$\frac{(n+1)z_2}{2(\sqrt{2\pi}\varpi)^3(2\pi)^2\epsilon_{\theta_1}\epsilon_{\theta_2}\epsilon_{\theta_3}(\max_{u_\lambda}-\min_{u_\lambda})\max_r}\exp(-\frac{1}{2\varpi^2}$$

$$u_a^2 + u_b^2 + u_c^2) I[\min_{u_i} \leq u_i \leq \max_{u_i}, \ , i = a, b, c], \tag{6.19}$$

where

$$z_2 = [\int_{\min u_c}^{\max_{u_c}} \int_{\min_{u_b}}^{\max_{u_b}} \int_{\min_{u_a}}^{\max_{u_a}} \frac{1}{(2\pi\varpi^2)^{\frac{3}{2}}} \exp -\frac{x^2 + y^2 + z^2}{2\varpi^2} dxdydz]^{-1}.$$

3. The Jacobian of the transformation from

$$(l_n^*, r, \Delta_1, \Delta_2, a_n^*, u_a, b_n^*, u_b, c_n^*, u_c, \theta_{1,n^*}, u_{\theta_1}, \theta_{2,n^*}, u_{\theta_2}, \theta_{3,n^*}, u_{\theta_3} \lambda_n^*, u_\lambda)$$

to

$$(l_n, l_{n+1}, a_n, b_n, c_n, a_{n+1}, b_{n+1}, c_{n+1}, \theta_{1,n}, \theta_{1,n}, \theta_{1,n}, \theta_{1,n+1},$$
$$\theta_{2,n+1}, \theta_{3,n+1}, \lambda_n, \lambda_{n+1})$$

using equations (6.13) through (6.16) is computed and is equal to

$$\frac{(\sin(\Delta_2))(|\cos(\Delta_1)| + 1)(|\sin(\Delta_1)| + 1)(|\cos(\Delta_2| + 1)}{1024r^2}. \tag{6.20}$$

4. The fourth ratio

The likelihood ratio will involve all the voxels that are currently occupied by x_n or x_{n+1} in addition to the voxels that will be covered by cell x_n^*. That is it involves $\Re(x_n) \cup \Re(x_{n+1}) \cup \Re(x_n^*)$, where $\Re(x_i)$ is the set of indices j such that the signals Y_j contains a contribution from at least one voxel in the object x_i. The likelihood ratio can be written as

$$\prod_{j \in \Re(x_n) \cup \Re(x_{n+1}) \cup \Re(x_n')} \frac{\sigma_{(j|x)}}{\sigma_{(j|x')}} \exp(-\frac{(y_j - \mu_{(j|x')})^2}{2\sigma_{(j|x')}^2} + \frac{(y_j - \mu_{(j|x)})^2}{2\sigma_{(j|x)}^2}). \tag{6.21}$$

Combining the four main ratios evaluated in equations (6.18), (6.21), (6.19) and (6.20) we get R. The acceptance probability for the merge move is $\min\{1, R\}$ and for the split move that will take the state from x' to x is $\min\{1, R^{-1}\}$.

Chapter 7

3D sampler performance

We have built up a model to fit the images obtained using confocal microscopy. In the previous chapter we have described the model for 3D images. The evaluation of this model has to be done using an MCMC simulation algorithm as we did for the 2D model. To set up values for the hyper-parameters involved in the model and in the proposal kernels of the algorithm we shall choose the same values as for the 2D problem where this is appropriate. For the distribution of the third axis of the ellipsoid, c, we set $\mu_c = 3.0$ and $\gamma_c^2 = 4.0$. Starting with an empty configuration, we ran the program for 50000 iterations. The resulting sample obtained at the final iteration is displayed in figure $(7-1)$. We viewed the sample by displaying three horizontal sections which correspond to stacks number 3, 7 and 12 arranged vertically from top to bottom as seen in figure $(7-2)$. In this thesis, the 2D configurations are displayed in a window where the vertical-axis represents the $x-$axis and the $y-$direction is the horizontal axis. The resulting sample has managed to locate some cells in the correct place but needs more iterations to adjust these cells and to locate the other cells in their correct places and positions.

Figure $(7-3)$ traces the cumulative total number of accepted moves with time recorded every 250 sweeps. Comparing these figures with the corresponding one of the 2D case displayed in figure $(3-13)$ we notice the decrease in the total number of the

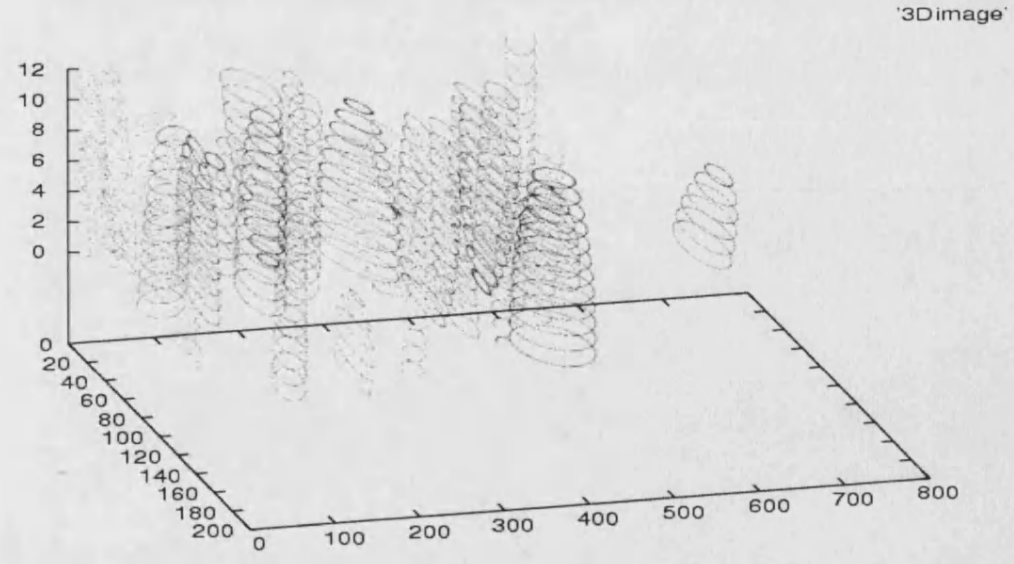

Figure 7-1: The resulting image in 3D viewed after 50000 iterations.

accepted moves for the same run time. The CPU time needed to implement a single iteration in the 3D problem is larger than that for the 2D problem because the data set is $k = 12$ times larger. Moreover, the rate of acceptance of the MCMC moves is lower due to the increase in the dimensionality of these moves. As an example, in the resize move in 2D there are generally 3^2 possible directions in which to propose the way of resizing a cell, while in 3D the number of possible directions increases to 3^3. Thereupon the probability of proposing good candidates for this move is lower in 3D than in 2D. The same problem occurs for the other move types. To decrease the time needed for convergence we can use more restricted prior distributions for the variables involved in the model depending on the data set. That is we can allow the data to supply the additional assumptions to define these parameter distributions rather than using uniform distributions. The question of what do data tell us about possibly many models that are justifiable from prior experience and the scientific context of the problem is considered by Aitkin (1991) in the model comparison choice problem when the parameter distributions are unknown. Aitkin used the whole data vector

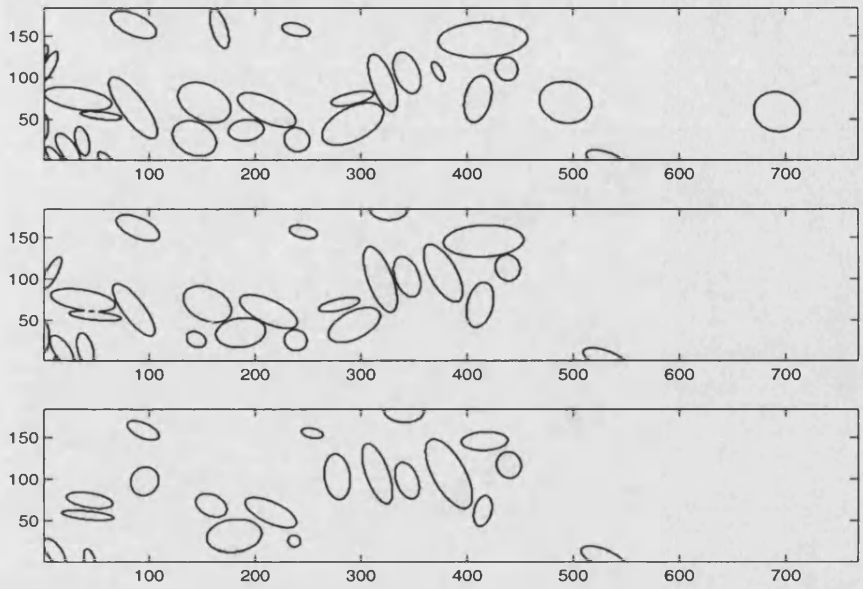

Figure 7-2: The resulting 2D sections orthogonal to the $z-$axis using z=3 (top), z=7 (middle) and z=12 (bottom) corresponds to the 50000 th iteration using the original 3D MCMC algorithm and starting from an empty configuration.

Y to supply additional assumptions about the unknown parameters of the posterior distributions and reuse the data to evaluate that posterior density. In addition to defining a restricted parameter distribution for the mark variables in the reference model, we suggest a careful off-line tuning of state dependent proposals for some moves in constructing the MCMC algorithm. These conditional proposals do depend on the present situation instead of proposing candidates that are completely random. So we try to be inventive in designing these proposals. These changes need to be done in a legal way to satisfy the detailed balance condition, thus the following adjustments have been made for some moves.

7.1 Improving the birth move

If the birth move is selected with probability $j_1(x)$, then we consider adding a new object say, x_{n+1}, to the current configuration, $x = \{x_1, \ldots, x_n\} \in \mathcal{U}^n$. Currently we randomly draw a new location $l_{n+1} = (c_{xn+1}, c_{y_{n+1}}, c_{zn+1})$ and then propose a mark

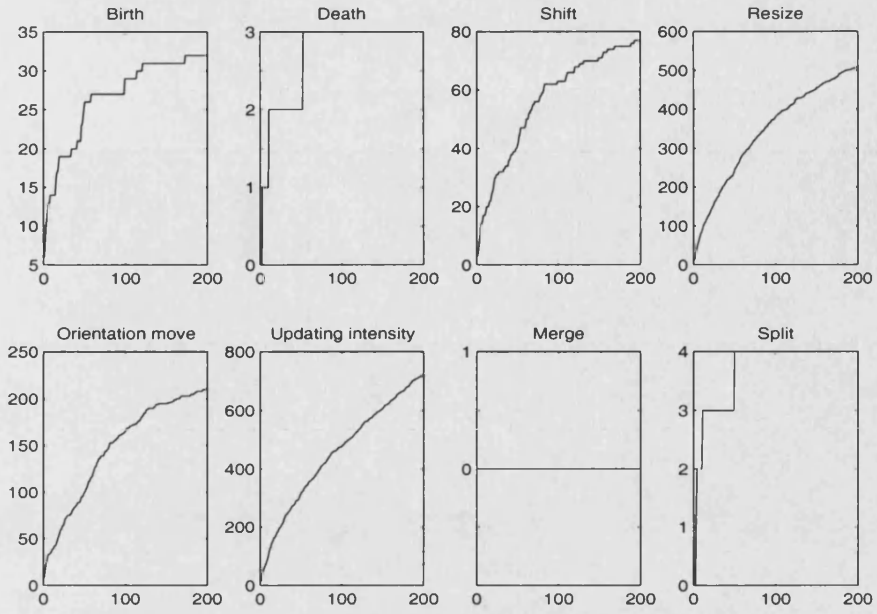

Figure 7-3: The trace of the cumulative total number of accepted moves recorded every 250 iterations applying the original 3D MCMC algorithm and starting from an empty configuration.

$m_{n+1} = (a_{n+1}, b_{n+1}, c_{n+1}, \theta_{1,n+1}, \theta_{2,n+1}, \theta_{3,n+1}, \lambda_{n+1})$ independently using the distributions defined in section (6.6) for each variable.

7.1.1 Modifying the proposal of a new location

In the birth moves we used to propose a new cell which does not overlap with the other cells by generating its centre location l uniformly over the window L and its mark vector m from its prior. If the cell does overlap with the other cells we set the acceptance probability of the move to zero, hence, the move is rejected. For the 3D problem, (c_x, c_y, c_z) represents the Cartesian coordinates of the location centre l of a cell and the mark m is a vector of $(a, b, c, \theta_1, \theta_2, \theta_3, \lambda)$. This proposal mechanism is modified so that it is not only a function of the record of each voxel ijk, $i = 1, \ldots, S_x, j = 1, \ldots, S_y, k = 1, \ldots, S_z$ alone but also of the average value of the records of voxels neighbouring voxel ijk, say ∂_{ijk}. That is, we use a smooth function $\bar{Y}_{ijk}$ for voxel ijk, where $\bar{Y}_{ijk}$ is the average of Y_l over $l \in \partial_{ijk}$. The neighbourhood structure ∂_{ijk} is local,

we chose ∂_{ijk} to be a cube of voxels with centre ijk and a side length of 5 voxels. If the voxel ijk is at the edge of the image we use the average of the available records in ∂_{ijk}. Using the smooth average value of each record $\bar{Y}_{ijk}$ is expected to decrease the chance of considering an isolated high value background voxel as the centre of a cell. Moreover, it will direct the potential centre of the new cell toward the mass of the cell where there are higher records and where there is less variation than at the edges. We assign to each voxel a probability associated with $\bar{Y}_{ijk}$. The defined density $p(ijk), i = 1,\ldots,S_x, j = 1,\ldots,S_y, k = 1,\ldots,S_z$ depends as well on k, the depth of voxel ijk. That is

$$p_{ijk} \propto \begin{cases} 1 & \bar{Y}_{ijk} < 70\Upsilon^{k-1} \\ \frac{\bar{Y}_{ijk}}{7\Upsilon^{k-1}} - 9 & 70\Upsilon^{k-1} \leq \bar{Y}_{ijk} < 210\Upsilon^{k-1} \\ 21 & 210\Upsilon^{k-1} \leq \bar{Y}_{ijk}, \end{cases}$$

where Υ is the parameter of the geometric attenuation. The above density ensures that for each optical section we assign the same probabilities to the corresponding record values after adjustment for the attenuation effect. There will be corresponding changes to the death proposals and a note should be made of the voxel ijk and its associated probability p_{ijk} for the computation of the acceptance probability for the birth move as well as the reverse death move.

7.1.2 Modifying the intensity proposal

Previously, we proposed the intensity level λ_{n+1} of the new cell uniformly over $M_\lambda = [\min_\lambda, \max_\lambda]$. Since the record of the voxel covering the centre of a cell can be used as an indicator of the intensity of that cell, we shall use the smooth data average of voxel $i\breve{j}k$, $\bar{Y}_{i\breve{j}k}$ to give a more representative intensity candidate for the new cell. Here $i\breve{j}k$ is the voxel that covers the new centre location l_{n+1}. The new candidate is still generated uniformly but over a more compact interval $[\min_{\lambda^*}, \max_{\lambda^*}]$, generally centred

at $\bar{Y}_{\check{ijk}} - \tau_0$ as we define $\min_{\lambda^*}$ to be

$$\max(\min_{\lambda}, \bar{y}_{\check{ijk}} - \tau_0 - \delta_\lambda),$$

and $\max_{\lambda^*}$ as

$$\min(\max_{\lambda}, \bar{y}_{\check{ijk}} - \tau_0 + \delta_\lambda),$$

where δ_λ is an arbitrary positive constant controlling the interval length.

Applying the above technique together with the one described in section (7.1.1) will induce changes in the acceptance probabilities for the birth and corresponding death moves. The resulting acceptance probability for the birth is $\alpha = \min\{1, R\}$, where

$$R = \frac{\beta(\max_{\lambda^*} - \min_{\lambda^*})}{(n+1)p_{\check{ijk}}(\max_{\lambda} - \min_{\lambda})} \prod_{ijk \in \Re(x_{n+1})} \frac{\sigma_{(ijk|x_{n+1}=\emptyset)}}{\sigma_{(ijk|x_{n+1})}}$$
$$\exp\{-\frac{(y_{ijk} - \mu_{(ijk|x_{n+1})})^2}{2\sigma^2_{(ijk|x_{n+1})}} + \frac{(y_{ijk} - \mu_{(ijk|x_{n+1}=\emptyset)})^2}{2\sigma^2_{(ijk|x_{n+1}=\emptyset)}}\} I[\text{no overlap}], \qquad (7.1)$$

where $\Re(x_{n+1})$ represents the set of indices j in which the signals Y_j are affected by cell x_{n+1}. For the overlapping condition we set $\alpha(x, x')$ to be zero if the proposed object overlaps with the existing ones within the boundaries of the window L. For the acceptance probability for the corresponding death move taking the chain from x' to x, we use $\min\{1, R^{-1}\}$ if $\lambda \in \bar{y}_{\check{ijk}} - \tau_0 - \delta_\lambda$ where ijk is the voxel containing the current centre of the cell x_{n+1}. The acceptance probability of the death move is set to zero if the cell could not appear at birth with this λ.

7.2 Improving the proposals for the shift move

Suppose that we want to update the location $l_t = (c_x, c_y, c_z)$ of a randomly selected cell x_t. We used to draw d_x, d_y and d_z from normal distributions with means $\nu_{d_j} =$

$0, j = x, y, z$ and standard deviations δ_x, δ_y and δ_z respectively, and set

$$c'_x = c_x + d_x \quad c'_y = c_y + d_y \quad c'_z = c_z + d_z.$$

The form of the cells is such that the voxels located in the central mass of the cells tend to have higher signals than the edge voxels. Hence, the proposal location should aim towards voxels with high records. Let ijk be the voxel containing $l_t = (c_x, c_y, c_z)$. We shall compare the value of the smooth average of voxel ijk, that is $\bar{Y}_{ijk}$ with the averages of some neighbouring voxels, and point the shift move in the direction of the highest average voxel. This technique is applied by controlling the means of the d_j, $j = x, y, z$, to be in the required directions. We shall compute each mean separately by observing its corresponding direction and fixing the others. For example in the direction of the $x-$axis, we compare the smooth average of voxels covering the points $(i + 2\delta_x)jk$, ijk and $(i - 2\delta_x)jk$, and set ν_{d_x} to be the distance and the direction between ijk and the voxel with the highest average value, i.e., $\nu_{d_x} \in \{-2\delta_x, 0, 2\delta_x\}$. Similarly we compute ν_{d_y} and ν_{d_z}. We then generate d_x, d_y, d_z from Normal distributions with means ν_{d_x}, ν_{d_y} and ν_{d_z} respectively and define the new location $l'_t = (c'_x, c'_y, c'_z)$ as before.

There is strong dependence between the intensity of a cell and the location of that cell. The signal of the centre point of a cell is a good estimate for the intensity level of that cell. Moreover, the variances of the cells are higher than the variance of the background areas, records with high value tend to be considered as part of cells rather than as background media in the resize move. Therefore we want to decrease the probability of considering high background records as part of cells by replacing the current intensity with a higher one directly when proposing the shift to a new location. We shall propose to update the intensity λ_t with a higher one within the same move.

Conditioning on the new candidate l'_t and the current intensity level λ_t, we shall generate λ'_t from a truncated normal distribution with mean $\eta_{\lambda'_t|\lambda_t,l'_t}$ and variance $\gamma^2_{\lambda'_t|\lambda_t,l'_t}$. Let $i\breve{j}k$ be the voxel covering the new location l'_i, then if the smooth average of $i\breve{j}k$ $\bar{Y}_{i\breve{j}k}$

is higher than λ_t, we set

$$\eta_{\lambda'_t|\lambda_t,l'_t} = \frac{\lambda_t + \bar{y}_{i\check{j}k}}{2} \qquad \gamma_{\lambda'_t|\lambda_t,l'_t} = \frac{\bar{y}_{i\check{j}k} - \eta_{\lambda'_t|\lambda_t,l'_t}}{2},$$

otherwise we set

$$\eta_{\lambda'_t|\lambda_t,l'_t} = \lambda_t \qquad \gamma_{\lambda'_t|\lambda_t,l'_t} = \frac{\eta_{\lambda'_t|\lambda_t,l'_t} - \bar{y}_{i\check{j}k}}{4},$$

where the normal distribution with these means and variances is restricted to the range $[\min_\lambda, \max_\lambda]$.

Suppose the current configuration is $x = \{x_1, \ldots, x_n\}$ and the shift move is selected, we randomly select an object say x_t and propose to update the location of its centre l_t to l'_t and its intensity λ_t to λ'_t. Let x'_t denote the resulting cell after updating x_i and let $x' = \{x_1, \ldots, x'_i, \ldots, x_n\}$, then the probability of acceptance of this move is defined to be

$$\alpha((l_t, \lambda_t), (l'_t, \lambda'_t)) = \min\left\{1, \frac{\pi(x', n)}{\pi(x, n)} \frac{\mathcal{L}(y|x')}{\mathcal{L}(y|x)} \frac{q((l'_t, \lambda'_t), (l_t, \lambda_t))}{q((l_t, \lambda_t), (l'_t, \lambda'_t))}\right\}. \tag{7.2}$$

The ratios on the right hand side of the above expression are equal to

$$\frac{a_1 a_3}{a_2 a_4} \frac{\gamma_{\lambda'_t|\lambda_t,l'_t}}{\gamma_{\lambda_t|\lambda'_t,l_t}} \exp[-\frac{(-d_x - \nu_{d_{x|l'_t}})^2}{2\delta_x^2} + \frac{(d_x - \nu_{d_{x|l_t}})^2}{2\delta_x^2} - \frac{(-d_y - \nu_{d_{y|l'_t}})^2}{2\delta_y^2}$$
$$+ \frac{(d_y - \nu_{d_{y|l_t}})^2}{2\delta_y^2} - \frac{(-d_z - \nu_{d_{z|l'_t}})^2}{2\delta_z{}^2} + \frac{(d_z - \nu_{d_{z|l_t}})^2}{2\delta_z^2} - \frac{(\lambda_t - \eta_{\lambda_t|\lambda'_t,l_t})^2}{2\gamma^2_{\lambda_t|\lambda'_t,l_t}} +$$
$$\frac{(\lambda'_t - \eta_{\lambda'_t|\lambda_t,l'_t})^2}{2\gamma^2_{\lambda'_t|\lambda_t,l'_t}}] \prod_{j \in \Re(x_t) \cup \Re(x'_t)} \frac{\sigma_{(j|x_t)}}{\sigma_{(j|x'_t)}} \exp(-\frac{(y_j - \mu_{(j|x'_t)})^2}{2\sigma^2_{(j|x'_t)}} + \frac{(y_j - \mu_{(j|x_t)})^2}{2\sigma^2_{(j|x_t)}})$$
$$I[\text{no overlap between the existing objects}],$$

which is simplified to

$$\frac{a_1 a_3}{a_2 a_4} \frac{\gamma_{\lambda'_t|\lambda_t,l'_t}}{\gamma_{\lambda_t|\lambda'_t,l_t}} \exp[-\frac{(\lambda_t - \eta_{\lambda_t|\lambda'_t,l_t})^2}{2\gamma^2_{\lambda_t|\lambda'_t,l_t}} + \frac{(\lambda'_t - \eta_{\lambda'_t|\lambda_t,l'_t})^2}{2\gamma^2_{\lambda'_t|\lambda_t,l'_t}}] \prod_{j \in \Re(x_t) \cup \Re(x'_t)} \frac{\sigma_{(j|x_t)}}{\sigma_{(j|x'_t)}}$$

$$\exp(-\frac{(y_j-\mu_{(j|x'_t)})^2}{2\sigma^2_{(j|x'_t)}}+\frac{(y_j-\mu_{(j|x_t)})^2}{2\sigma^2_{(j|x_t)}})I[\text{no overlap}],$$

where $\mu_{(j|x_t)}$ and $\sigma^2_{(j|x_t)}$ are the attenuated mean and variance of the record at voxel record j respectively conditioned on x_t, and $\mu_{(j|x'_t)}$ and $\sigma^2_{(j|x'_t)}$ are the attenuated mean and variance of the record of voxel j respectively given x'_t. The values of $\{a_1, a_2, a_3, a_4\}$ are normalising constants and are equal to

$$a_1=[\int_{-c_{x'_t}}^{L_x-c_{x'_t}}\int_{-c_{y'_t}}^{L_y-c_{y'_t}}\int_{-c_{z'_t}}^{L_z-c_{z'_t}}\frac{1}{(2\pi)^{\frac{3}{2}}\delta_x\delta_y\delta_z}\exp-\frac{1}{2}(\frac{(w-\nu_{d_{x|l'_t}})^2}{\delta_x^2}+\frac{(y-\nu_{d_{y|l'_t}})^2}{\delta_y^2}+\frac{(z-\nu_{d_{z|l'_t}})^2}{\delta_z^2})dzdydw]^{-1}$$

$$a_2=[\int_{-c_{xt}}^{L_x-c_{xt}}\int_{-c_{y_t}}^{L_y-c_{y_t}}\int_{-c_{zt}}^{L_z-c_{zt}}\frac{1}{(2\pi)^{\frac{3}{2}}\delta_x\delta_y\delta_z}\exp-\frac{1}{2}(\frac{(w-\nu_{d_{x|l_t}})^2}{\delta_x^2}\frac{(y-\nu_{d_{y|l_t}})^2}{\delta_y^2}\frac{(z-\nu_{d_{z|l_t}})^2}{\delta_z^2})dzdydw]^{-1}$$

$$a_3=\left[\Phi(\frac{\max\lambda-\eta_{\lambda_t|\lambda'_t,l_t}}{\gamma_{\lambda_t|\lambda'_t,l_t}})-\Phi(\frac{\min\lambda-\eta_{\lambda_t|\lambda'_t,l_t}}{\gamma_{\lambda_t|\lambda'_t,l_t}})\right]^{-1}$$

$$a_4=\left[\Phi(\frac{\max\lambda-\eta_{\lambda'_t|\lambda_t,l'_t}}{\gamma_{\lambda'_t|\lambda_t,l'_t}})-\Phi(\frac{\min\lambda-\eta_{\lambda'_t|\lambda_t,l'_t}}{\gamma_{\lambda'_t|\lambda_t,l'_t}})\right]^{-1}.$$

7.3 Improving the rotations

7.3.1 Changing the prior distributions for the angles of rotations

Observing figures $(6-6)$, $(6-7)$ and $(6-8)$ we realize that the rotations of the objects in the window L are significant in the $xy-$plane and are not apparent in the planes involving the $z-$axis. In other words θ_2 is close to zero or to π. To reflect this situation we make a structural modification that concerns the choice of the prior of the angle θ_2. Instead of the prior distribution of θ_2 defined in section (6.6), an alternative is to

use a prior that assigns higher probabilities to values close to zero or to π and smaller probabilities for the other values of θ_2 than the previous density. We define

$$\pi_{\theta_2}(\theta_2) = \begin{cases} \frac{b}{2c_{\theta_2}(b-1)+\pi} & \text{when } \theta_2 \leq c_{\theta_2} \text{ or } \theta_2 \geq \pi - c_{\theta_2} \\ \frac{1}{2c_{\theta_2}(b-1+\pi} & \text{when } c_{\theta_2} < \theta_2 < \pi - c_{\theta_2} \end{cases}, \tag{7.3}$$

where b and c_{θ_2} are arbitrary positive constants, $0 < c_{\theta_2} < \pi - c_{\theta_2}$. Initially we set b to 20. Picking small values for c_{θ_2} will increase the probability of having θ_2 close to zero or to π. If this is the case, then the rotations of the cells are effectively in the xy plane only. This density is symmetric about $\pi/2$ and is able to accommodate the required values of θ_2.

If θ_2 is equal to zero then the angle of rotation of the $xy-$plane to the $x'y'-$plane is equal to $\theta_1 + \theta_3$, whereas if θ_2 is equal to π then the angle of rotation of the $xy-$plane to the $x'y'-$plane is equal to $\theta_1 - \theta_3$. Figure $(6-6)$ indicates that the required rotations of the $x-$axis and $y-$axis to the new $x'-$axis and $y'-$axis are mainly through angles close to zero or π. Hence, we shall try to control the rotation of the $xy-$plane to the new $x'y'-$plane by changing the values of θ_1 more freely and assigning values to θ_3 close to zero or to π. So for θ_3 we define the same density as defined for θ_2 in equation (7.3).

Changing the prior distributions for the angles of rotations θ_i, $i = 2, 3$ will affect the birth move, however, the acceptance rate will remain as it was as long as we use the same densities for generating the angles of rotation for the new cell as their priors. The changes in the priors of the angles of rotation will affect acceptance rate for the orientation move, and the acceptance rates for the split and merge moves. For the orientation move, the acceptance probability for changing $\underline{\theta_t} = (\theta_{1,t} \quad \theta_{2,t} \quad \theta_{3,t})$ to $\underline{\theta_t}' = (\theta'_{1,t} \quad \theta'_{2,t} \quad \theta'_{3,t})$ for a randomly selected object t from the current existing

configuration $x = \{x_1, \ldots, x_n\}$ will change to

$$\alpha(\underline{\theta_t}, \underline{\theta_t}') = \min\{1, \frac{|\cos(\theta'_{1,t})| + \frac{1}{\pi}}{|\cos(\theta_{1,t})| + \frac{1}{\pi}} \frac{\pi_{\theta_2}(\theta'_{2,t})}{\pi_{\theta_2}(\theta_{2,t})} \frac{\pi_{\theta_2}(\theta'_{3,t})}{\pi_{\theta_2}(\theta_{3,t})} \prod_{j \in \Re(x_t) \cup \Re(x'_t)} \frac{\sigma_{(j|x_t)}}{\sigma_{(j|x'_t)}}$$
$$\exp(-\frac{(y_j - \mu_{(j|x'_t)})^2}{2\sigma^2_{(j|x'_t)}} + \frac{(y_j - \mu_{(j|x_t)})^2}{2\sigma^2_{(j|x_t)}}) I[\text{no overlap}]. \quad (7.4)$$

(This is when using the same proposal kernels for the candidate angles as defined in section (3.5.4) for the 2D problem after generalising it to handle the three angles of rotation). The prior densities for the angles of rotation are also involved in the calculation of the acceptance probabilities of the merge and split moves. For the merge move defined in section (6.7.1), equation (6.18) will change to

$$\frac{z_1^{-1}(3\pi)(2\pi)^{\frac{3}{2}}\gamma^2\gamma_c(\pi|\cos(\theta^*_{1,n})| + 1)\pi_{\theta_2}(\theta^*_{2,n})\pi_{\theta_3}(\theta^*_{3,n})(\max_\lambda - \min_\lambda)}{\beta(\pi|\cos(\theta_{1,n})| + 1)(\pi|\cos(\theta_{1,n+1})| + 1)\pi_{\theta_2}(\theta_{2,n})\pi_{\theta_2}(\theta_{2,n+1})\pi_{\theta_3}(\theta_{3,n})\pi_{\theta_3}(\theta_{3,n+1})}$$
$$\exp -\frac{1}{2\gamma^2}((a^*_n - \mu_a)^2 + (b^*_n - \mu_b)^2 + (c^*_n - \mu_c)^2 - (a_n - \mu_a)^2 - (b_n - \mu_b)^2$$
$$-(c_n - \mu_c)^2 - (a_{n+1} - \mu_a)^2 - (b_{n+1} - \mu_b)^2) - (c_{n+1} - \mu_c)^2)$$
$$I[\text{no overlap}] I[\min_{axis} \leq a^*_n, a_n, a_{n+1}, b^*_n, b_n, b_{n+1}, c^*_n, c_n, c_{n+1} \leq \max_{axis}], \quad (7.5)$$

7.3.2 Modifying the proposals for the orientation move

Although we expect changing the prior densities of the angles of rotation to produce angles more representative of the true angles, we shall also modify the proposal mechanism of the orientation to help further in controlling and directing the angles of rotations of the cells to the required values. The proposal kernel of θ_1 will be the same as defined earlier; that is, we propose a new candidate θ'_1 uniformly over $(\theta_{1,t} - \eta, \theta_{1,t} + \eta)$ for some fixed value η. To help to direct the proposal candidates $\theta'_{2,t}$ and $\theta'_{3,t}$ towards zero or towards π we shall define two random variables d_{θ_2} and d_{θ_3}. We shall draw them from two independent normal distributions with means $\mu_{d_{\theta_j}}$, $j = 2, 3$ and variance σ^2_θ, where

$$\mu_{d_{\theta_j}} = \begin{cases} -\theta_j & \text{if } \theta_j < \frac{\pi}{2}, \\ \pi - \theta_j & \text{if } \theta_j \geq \frac{\pi}{2} \end{cases}$$

for $j = 2$ and 3, and we set $\sigma_\theta = \frac{\pi}{20}$. We then set $\theta'_{j,t} = (\theta_{j,t} + d_{\theta_j})$, $j = 2, 3$. If $\theta'_{j,t} < 0$ then we set $\theta'_{j,t} = \theta'_{j,t} + \pi$ and if $\theta'_{j,t} > \pi$ then we set $\theta'_{j,t} = \theta'_{j,t} - \pi$. The acceptance rate when applying these proposal kernels for the orientation move is

$$\alpha(\underline{\theta_t}, \underline{\theta_t}') = \min\{1, \frac{(|\cos(\theta'_{1,i})| + \frac{1}{\pi})}{(|\cos(\theta_{1,i})| + \frac{1}{\pi})} \frac{\pi_{\theta_2}(\theta'_2)}{\pi_{\theta_2}(\theta_2)} \frac{\pi_{\theta_3}(\theta'_3)}{\pi_{\theta_3}(\theta_3)}$$
$$\exp[-\frac{1}{2\sigma_\theta^2}(-d_{\theta_2} - \mu_{d_{\theta'_2}})^2 - (d_{\theta_2} - \mu_{d_{\theta_2}})^2 + (-d_{\theta_3} - \mu_{d_{\theta'_3}})^2 - (d_{\theta_3} -$$
$$\mu_{d_{\theta_3}})^2] \prod_{j \in \Re(x_t) \cup \Re(x'_t)} \frac{\sigma_{(j|x_t)}}{\sigma_{(j|x'_t)}} \exp(-\frac{(y_j - \mu_{(j|x'_t)})^2}{2\sigma^2_{(j|x'_t)}} + \frac{(y_j - \mu_{(j|x_t)})^2}{2\sigma^2_{(j|x_t)}})$$
$$I[\text{no overlap between the objects}]\}. \qquad (7.6)$$

7.4 Improving the merge and the split moves

7.4.1 Changing the distribution of Δ_2

Observing figure $(7-2)$ we notice the need to split some cells in the $xy-$plane along the $z-$axis such as cells occupying the grids $[x = 150 : 180, y = 80 : 110, z = 2 : 11]$, $[50 : 150, 370 : 410, 3 : 12]$ and $[100 : 160, 350 : 450, 1 : 12]$. To encourage splitting of the cells vertically along the $z-$axis, we shall control the spherical coordinate Δ_2 defined in section (6.7.1). If we want to split a cell x_C into two cells x_A and x_B along the $z-$axis we have to choose the centres of the two split cells in the same $xy-$plane by fixing the $z-$coordinates of l_A and l_B at approximately the same value. This is done by choosing Δ_2 close to $\pi/2$ or close to $3\pi/2$. Therefore instead of defining a uniform distribution over $(0, 2\pi)$ for Δ_2 we defined the following distribution:

$$\pi_{\Delta_2}(\Delta_2) = \begin{cases} \frac{20}{2\pi+76c_{\Delta_2}} & \text{if} \frac{\pi}{2} - c_{\Delta_2} \leq \Delta_2 \leq \frac{\pi}{2} + c_{\Delta_2} \text{and } \frac{3\pi}{2} - c_{\Delta_2} \leq \Delta_2 \leq \frac{3\pi}{2} + c_{\Delta_2} \\ \frac{1}{2\pi+76c_{\Delta_2}} & \text{otherwise} \end{cases}, \quad (7.7)$$

where c_{Δ_2} is an arbitrary constant, $0 < c_{\Delta_2} < \frac{\pi}{2}$.

7.4.2 Changing the distribution of r

Most of the attempted split moves are rejected because of the overlapping constraint, i.e., the two new cells overlap one another. To overcome this difficulty an adequate distance between the centres of the new cells $2r$ should be used. Small values of r result in having two very close cells so the probability of overlap of these cells will be high. On the other hand large r will increase the chance of the new cell overlapping with the cells in the neighbourhood of the original cell. In section (6.7.1) we drew r uniformly over $(0, \max_r)$, where the upper bound $\max_r$ is conditioned on Δ_1, Δ_2 and ϵ_d, and is computed using table (6.1). Although the lower bound of this interval is zero, practically it is bounded by $\min_{axis}$ because the cells have to be at least $2\min_{axis}$ apart not to overlap. We shall change the distribution of r to give higher probability for values of r that fall in the mid range of the valid interval. We use a normal distribution with mean μ_r where

$$\mu_r = \frac{(a_C|\cos(\Delta_1)| + b_C|\sin(\Delta_1)|)|\sin(\Delta_2)|}{2},$$

and variance σ_r^2 where

$$\sigma_r = \frac{|\max_r - \mu_r|}{3},$$

where $r \leq \max_r$, i.e., this normal distribution is restricted to the range $(0, \max_r)$. Here a_C and b_C are the semi-axes of the cell which it is proposed to split. The suggested changes for the distributions of Δ_2 and r will affect the proposal ratio involved in the calculation of the acceptance probability for the merge move and accordingly the

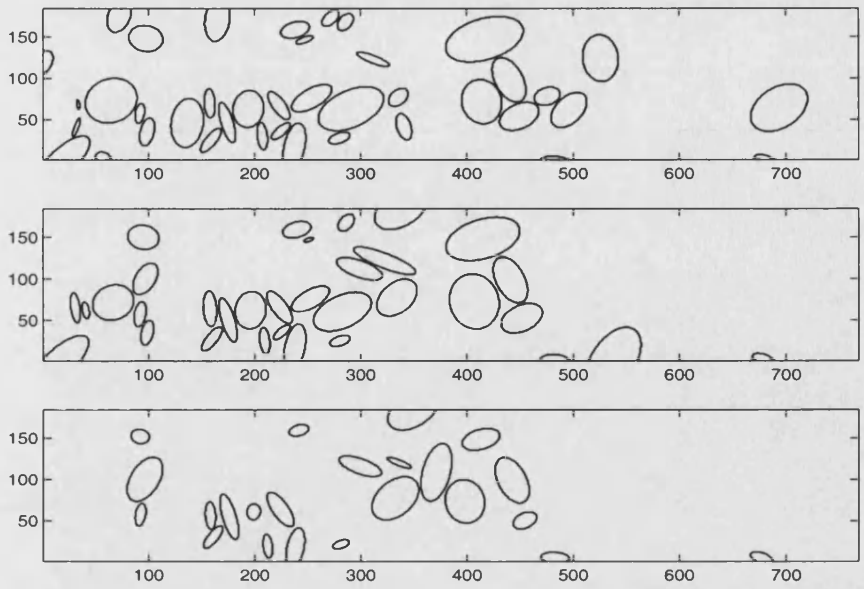

Figure 7-4: The resulting 2D sections orthogonal to the $z-$axis for z=3(top), 7(middle) and 12(bottom) corresponds to the 50000 th iteration after modifying the sampler and starting from a blank configuration.

split move. Suppose that the current configuration is $x = \{x_1, \ldots, x_{n-1}, x_n, x_{n+1}\}$ and we propose to combine two randomly selected cells say x_n, x_{n+1} to x_n^*. Let x' be $\{x_1, \ldots, x_{n-1}, x_n^*\}$ then the proposal ratio defined in equation (6.19) will become

$$\frac{(n+1)z_2\pi_{\Delta_2}(\Delta_2)z_r}{2(2\pi)^3\varpi^3\sigma_r\epsilon_{\theta_1}\epsilon_{\theta_2}\epsilon_{\theta_3}(\max_{u_\lambda}-\min_{u_\lambda})}\exp(-\frac{u_a^2+u_b^2+u_c^2}{2\varpi^2}-\frac{(r-\mu_r)^2}{2\sigma_r^2})$$
$$I[0<r<\max_r]I[\min_{u_a}\leq u_a\leq\max_{u_a},\min_{u_b}\leq u_b\leq\max_{u_b},\min_{u_c}\leq u_c\leq\max_{u_c}], \qquad (7.8)$$

where

$$z_r=[\int_0^{\max_r}\frac{1}{\sqrt{2\pi}\sigma_r}\exp-\frac{(w-\mu_r)^2}{2\sigma_r^2}dw]^{-1}.$$

7.5 Results

To test the performance of our sampler after applying the modifications, we set c_{θ_2} to $\pi/8$. Again the sampler had a run of 50000 iterations. Comparing the resulting output shown in figure $(7-4)$ with the input image displayed in figure (6-1) we noticed

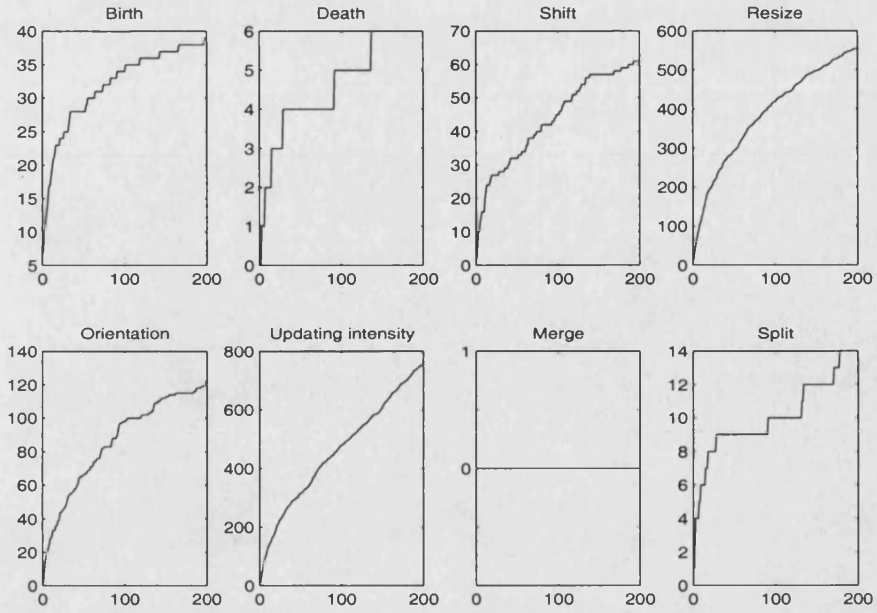

Figure 7-5: The trace of the cumulative number of accepted moves recorded every 250 iterations after modifying the sampler and starting from a blank state.

that there are several small adjacent cells occupying the region of big cells. Figure $(7-5)$ traces the total number of accepted moves during the run time recorded every 250 iterations. Comparing these results with the corresponding ones obtained before implementing the adjustments described in section (7.1) through section (7.4), the following changes have been noticed: The acceptance rate of the birth moves after the adjustments is indeed higher than it was before, which indicates the significant effect of the modified algorithm in implementing the birth move. The death moves are accepted more often than in the original sampler. This is due to the increase in the rate of acceptance of the split moves; the accepted death moves often involve cells resulting from accepted split moves. The shift move is accepted less often in the modified sampler. This was expected because our goal from combining two changes in this move; changing the location of the cell as well as changing its intensity, was to produce a better quality move rather than to increase the acceptance rate of the move. Finally a large decrease in the total number of accepted orientation moves is noticed. This may be as a result

of having more representative angles using the new prior densities for angles θ_2 and θ_3 (see figure $(7-4)$) or as a result of having very close cells in the scene which implies rejection of the proposed orientations due to overlapping.

For further investigation of the effects of the previous modifications, we will decrease the number of birth moves in order to give a better chance for the existing cells to improve their positions once created before adding new cells in the neighbourhood of these cells. Another run was conducted reducing the probability of picking the birth move. This resulted in good output samples which indicate an improvement. One of these samples is shown in figure $(7-6)$. Although the output samples obtained were final results after applying all the previous modifications, some of these modifications were more effective than the others. Decreasing the probability of choosing the birth move by one fourth allows cells to adjust themselves before adding new adjacent cells. We think that the changes in the shift and the orientation moves did lead to good results in the sense that it directed the proposal candidate to the mode of the distribution. On the other hand they limit the chance of exploring the sample space as they propose restricted candidates.

7.6 MCMC algorithm using a good starting state

It is apparent that the 3D samples obtained using our simulation technique for the target model and starting with the empty configuration take an enormously long time to reach convergence. One objective of the biologist in studying these data is to simulate a representative configuration of the image, in order to give a good summary for the cells in the specimen under study in practical time. To achieve this aim taking into consideration the limited resources and the slow mixing sampler we shall choose a starting state which is more informative than an extreme one such as an empty configuration.

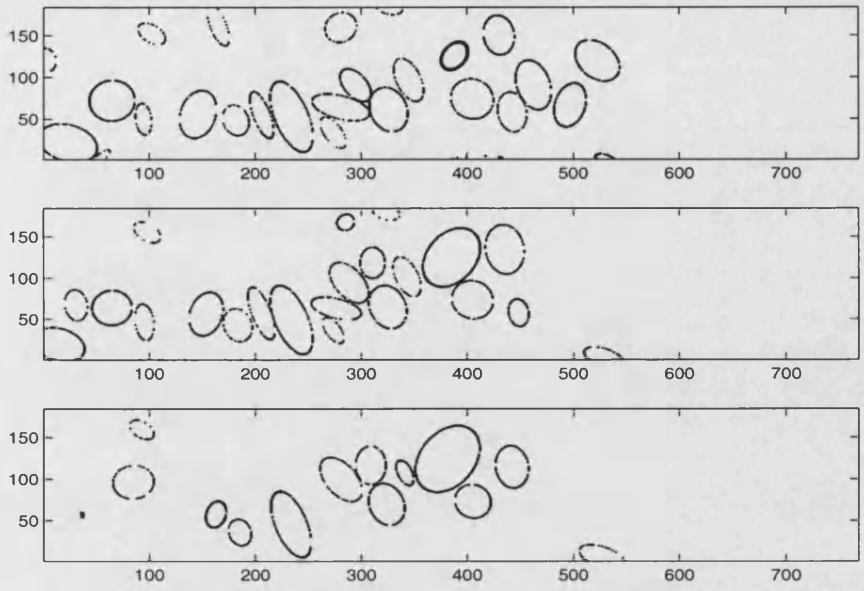

Figure 7-6: The resulting 2D sections orthogonal to the $z-$axis for $z = 3$ (top), $z = 7$ (middle) and $z = 12$ (bottom) corresponds to the 50000 th iteration after modifying the sampler and reducing the probability of picking the birth move.

7.6.1 Mathematical morphology for the 3D space

To construct a good starting state for our sampler, we shall generalise the algorithm developed in chapter (4) for the 2D case to handle the 3D space. In the first stage of the algorithm, we start at the 3D voxel level and we threshold the data to obtain a binary image. The thresholding process is applied taking into account the existence of the attenuation effects in the 3D data records. We set the threshold parameter to 90. This value is higher than the mean or the median of the first 2D optical section in the 3D set. Therefore the starting cells in the initial state would have high intensity levels. That is there is a degree of certainty about the existence of the cells. Then we use some mathematical morphology operators such as the opening operator to obtain several connected sets of voxels. The structuring element, B, to be used in the morphology will be a cube of $3 \times 3 \times 3$ voxels and with reference point being the centre of that cube. The resulting 3D output after applying the morphology opening operators is viewed using 3 horizontal sections corresponding to depth $3, 7$ and 12 respectively in figure

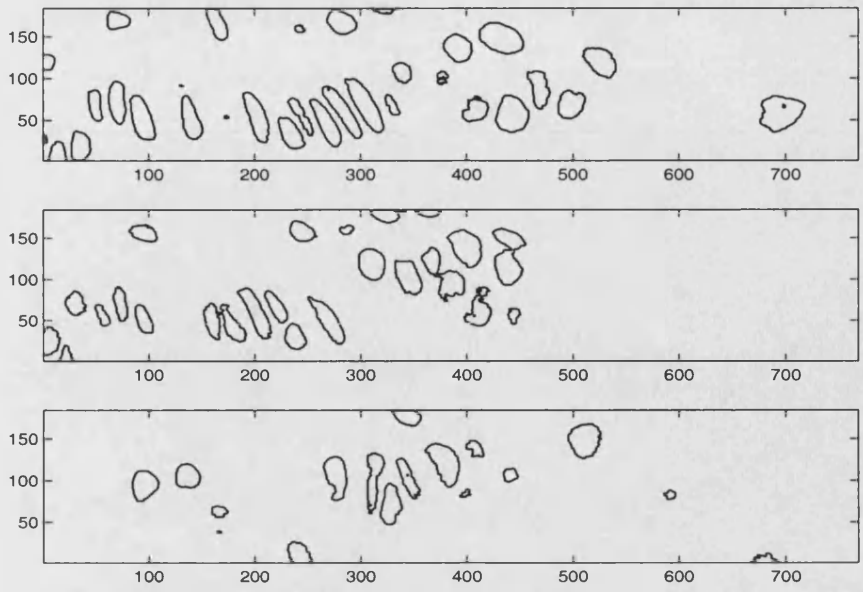

Figure 7-7: The resulting 2D sections at depths z=3, 7 and 12 sectioning the 3D output after applying morphological operators.

$(7-7)$.

7.6.2 Maximum variance concept

In the second stage of the algorithm after applying the opening operator, we convert the resulting connected voxels into ellipsoids. In chapter 4 we considered the 2D connected pixels as a level set of the bivariate normal distribution with X_1-variables being the row index of the grid and the X_2-variables being the column index. In this section we shall generalise this concept to the 3D problem using a $p-$variate normal density function where $p = 3$. We shall approach this problem by developing matrix methods that enable us to do so.

The p-variate normal distribution for p-variables $x_1, \ldots, x_p$ is

$$f(x_1 \ldots x_p) = (2\pi)^{-\frac{p}{2}} |\Sigma|^{-\frac{1}{2}} \exp(-\frac{\mathcal{X}^T \Sigma^{-1} \mathcal{X}}{2}),$$

where $\mathcal{X}^T = [x_1 - \mu_1, \ldots, x_p - \mu_p]$ and Σ is the variance-covariance matrix of the p-

variables. In the 3-variate normal distribution, we associate a 3 dimensional vector $[X_1, X_2, X_3]$ with a point in the 3-dimensional indexes of the 3D grid (i, j, k), representing the row, column and the depth index of a voxel.

We learn about the required ellipsoid forming the connected voxels using the following theorem:

Theorem 1 *Given $p-$variables following a $p-$variate normal distribution, the axis defining the linear transformed variable with the maximum variance is the major axis of the iso-density hyper-ellipsoid*

$$\mathcal{X}^T \Sigma^{-1} \mathcal{X} = C \tag{7.9}$$

where C is a arbitrary positive constant.

That is, $\mathcal{X}^T \Sigma^{-1} \mathcal{X} = C$ is the equation of the surface of the hyper-ellipsoid. For $p = 3-$variate case the resulting ellipsoid is identified using its main components: the centre, the three semi-axes and the angles of rotations. The centre of the ellipsoid is the mean vector of its $p = 3$ variables, $\mu^T = [\mu_{X_1}, \mu_{X_2}, \mu_{X_3}]$ and is estimated using the averages of X_1, X_2 and X_3. The semi-axes of the ellipsoid and the rotation angles as stated by the above theorem can be found by constructing a linear combination

$$Y^T = \mathcal{X}^T V + \mu^T,$$

such that the variance of Y

$$V^T \Sigma V, \tag{7.10}$$

will be as large as possible under the restriction that $V^T V = I$. Here $V = [V_1, V_2, V_3]$ is a matrix made up of 3×1 column vectors V_1, V_2, V_3. The restriction is meant to fulfill the orthogonality of the defined linear constructions. To solve the above equation we use the method of *Lagrange multipliers*, which defines an unknown variable λ called

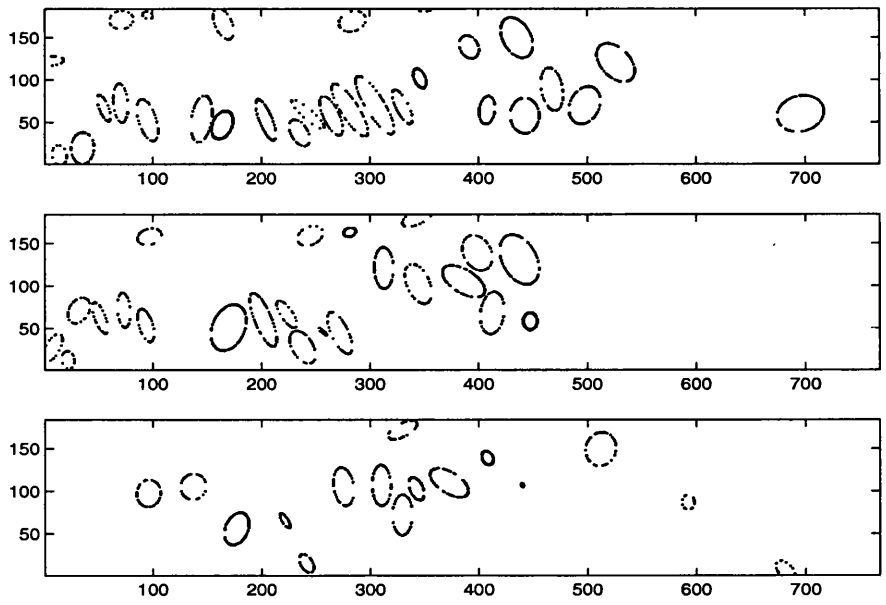

Figure 7-8: The resulting 3D configuration after converting the results from morphology operators into ellipsoids. The 3D state is viewed using three optical sections using z=3, 7 and 12.

the Lagrange multiplier and solves equation (7.10) by maximising

$$F = V^T \Sigma V - \lambda(V^T V - I).$$

To maximise F we take the partial derivative of F with respect to V and set the result to zero. The 3 possible roots λ_1, λ_2 and λ_3 for the above equation are called the eigenvalues. They are computed using the characteristic equation

$$|V - \lambda I| = 0.$$

We then solve for

$$(\Sigma - \lambda_i I)V_i = [0,0,0]^T, \quad i = 1,2,3. \tag{7.11}$$

substituting for the eigenvalues $\lambda_i, i = 1,2,3$ to get the corresponding eigenvectors of Σ, $V_i, i = 1,2,3$. By solving the eigenvalue problem we obtain the principle axes of the ellipsoid. The first principle axis with the largest λ corresponds to the major

axis (largest diameter), the second principle axis corresponds to the largest diameter orthogonal to the first. The third axis is the one that is orthogonal to both first and second axes. The lengths of the semi-axes are proportional to the square roots of the variances of Y in each direction. Using equation (7.11) and applying some matrix computations we can derive that the variances of the transformed variables Y are the eigenvalues themselves. To avoid overlapping between the resulting ellipsoids we define the length of the semi-axes to be 1.5 times the estimated standard deviations of the transformed variables bounded by the range of the axes space, $[\min_{axis}, \max_{axis}]$. If there are still overlapping cells then we set the proportion scale to a lower value. We extract the angles of rotations θ_1, θ_2 and θ_3 from the V matrix which corresponds to the transformed set of weights defining the rotation of a point in the 3D space as defined in detail in section (6.4). Finally we estimate the initial intensity of the cells using the average records inside each ellipsoid.

7.6.3 Results

Applying the above steps we get the 3D configuration presented in figure $(7-8)$. Using this 3D state as an initial configuration for our chain we run the MCMC sampler for 50000 iterations. The sample corresponding to the final iteration is shown in figure $(7-9)$. Figure $(7-10)$ traces the cumulative total number of accepted moves for that run updated every 500 iterations. The reason for not having any deaths during the run time is that the initial cells are added with a high degree of certainty using a high value for threshold parameter. In the resulting image we see new cells which did not exist in the starting configuration such as the cells located in the window $[1:20] \times [660:720] \times [10:12]$ and $[1:30] \times [400:600] \times [1:12]$. The positions of some cells have changed for a better fit to the data, for example with cells occupying the window $[140:160] \times [80:110] \times [5:8]$ and $[140:160] \times [230:260] \times [5:8]$. There were some efficient split moves such as the split of the cell covering the 3D region $[20:80, 150:200, 5:8]$ and $[80:140, 350:400, 5:8]$. These results demonstrate the

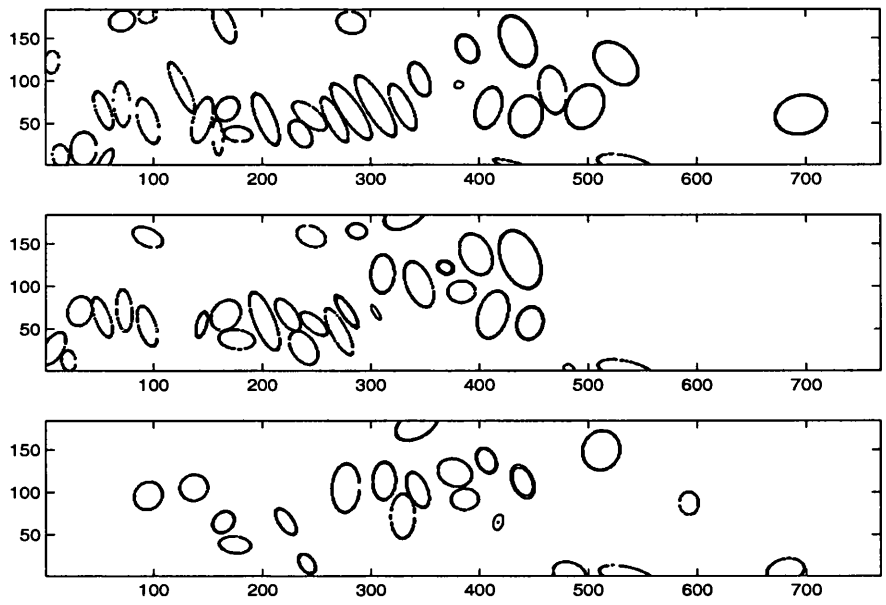

Figure 7-9: The resulting 3D sample corresponds to the 50000th iteration after running the MCMC sampler using the initial state. The 3D state is displayed using three optical sections orthogonal to z=3, 7 and 12.

need for a lengthy burn-in time, even if we start with an initial value near the mode of the target model. There is also uncertainty about some cells such as those occupying the region $[20:100, 110:200, 1:9]$. The voxels lying in that region have high record values and the cells have no sharp discernable edges so the background areas cannot be distinguished easily.

In conclusion, using a carefully selected initial value for the MCMC chain when the space of the configuration is very big certainly has positive effects on producing samples located near the mode of the target distribution.

7.7 Intervention of the user to construct a good starting state

To allow for the experience of the user in constructing a starting point for the chain, we shall apply some statistical tools to help him with his decision and to reduce the

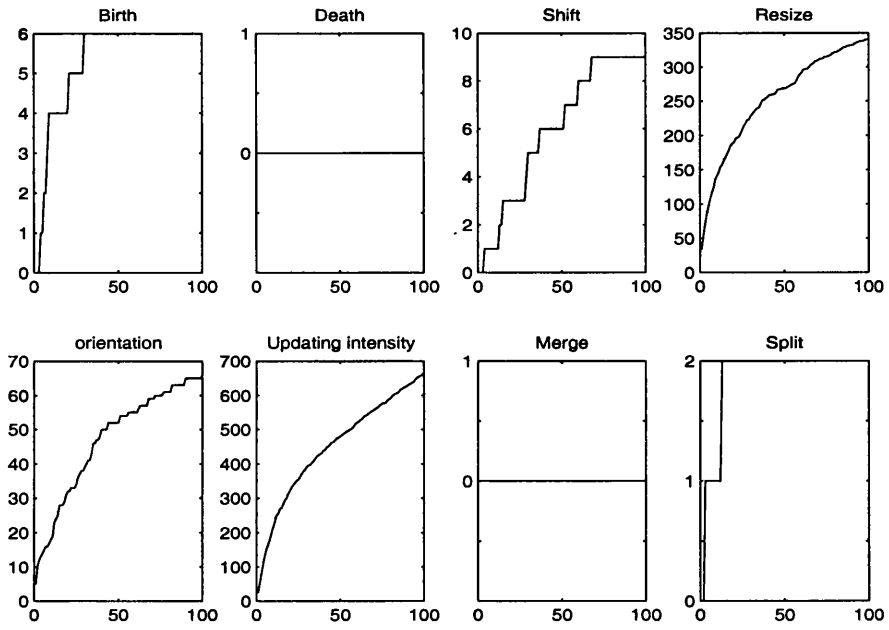

Figure 7-10: The trace of the total number of accepted moves recorded every 500 iterations for 50000 iterations MCMC run and starting from the state shown is figure 7-8.

time needed for his contribution. We shall start by applying the automated algorithm described in the preceding section, and then run the MCMC algorithm for some time. The user intervention is allowed after that stage; the user is asked to point out a single point inside each cell in the image about which he is certain. The points provided, p, are then checked and compared with the existing cells in the current configuration. Adjustments are carried out on these cells based on the information provided. This is achieved automatically using a special self consistent program taking into consideration the non overlapping condition. For illustration suppose a reference point is not covered by any cell in the sample, then we create an extra cell with centre at that reference point and assigning the lower bound to its the semi-axes a, b, c and zeros to the angles of rotation θ_1, θ_2 and θ_3 and $\min_\lambda$ to λ, the intensity level of the cell. This is done in the absence of any other information regarding the cell.

To get a similar reconstruction of the image without the above algorithm to help would require a lot more work for the user. The first part of the automated algorithm

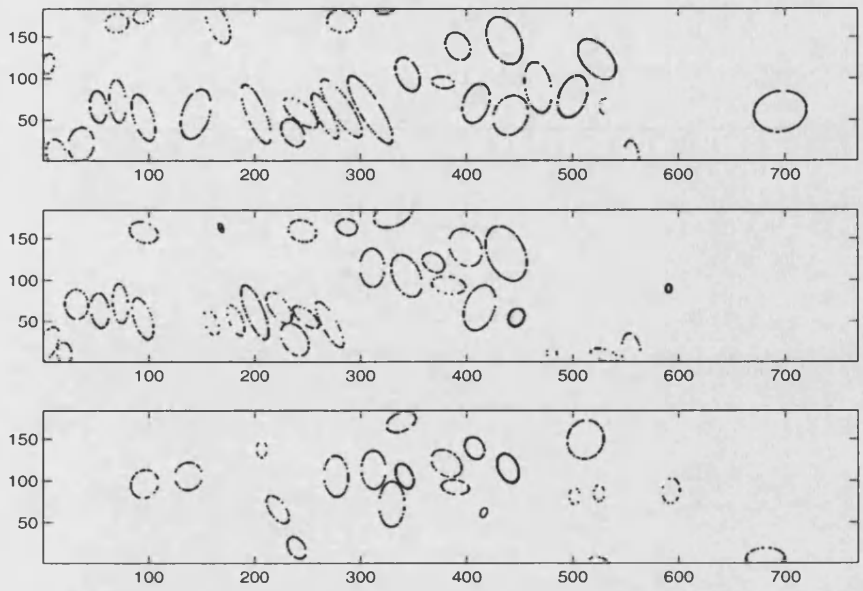

Figure 7-11: The 3D initial configuration after running the MCMC sampler for 50000 iterations and allowing user intervention. The 3D configuration is viewed using 2D optical sections corresponds (from top to bottom) to $z = 3, 7$ and 10.

finds out the cells, so the users do not have to click on them. Even for the missing cells, just one click in the centre is needed. These simple efforts are needed from the user to identify the cells rather than a lot of work outlining the cell surface in the 3D manually.

Since we are using expert knowledge along with some morphology operators depending on the data set and applying the MCMC algorithm, we then consider the configuration at that stage to be highly reliable. That is, we shall use the reference points in modifying our equilibrium distribution and in defining the MCMC sampler. All the configurations shall be defined on the same dimension space thus we shall use the standard Metropolis-Hastings algorithm to update the Markov chain. We shall use the conditional MCMC algorithm allowing only the moves that preserve the number of the cells such as shift, resize, orientation and updating the intensity level move. The proposed moves are only accepted if they ensure the existence of the reference point in the potential candidate cell.

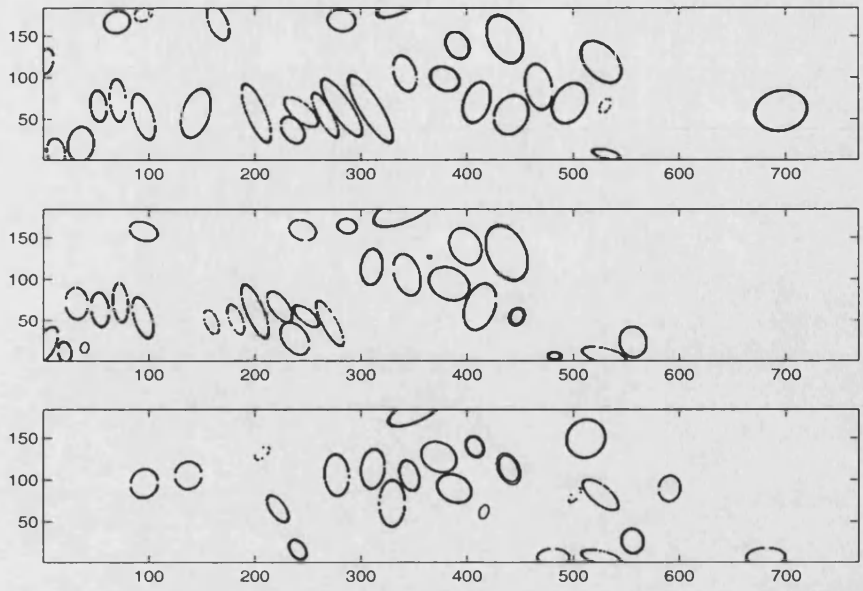

Figure 7-12: The resulting 2D optical sections at depths z =3, 7 and 12 corresponds to the 50000th iteration running the conditional MCMC algorithm and using the initial state which agrees with user knowledge.

To apply the above technique we use the 3D output sample resulting from the unconditional MCMC algorithm viewed in figure $(7-9)$. This configuration is used as an input image for the expert to provide the reference points. Adjustments on this sample state with regards to the reference points is done as described above and the resulting image is displayed in figure $(7-11)$. A run of 50000 iterations is then carried out using the conditional 3D MCMC sampler and the final sample is displayed in figure $(7-12)$. As expected the output sample is promising. The resulting output is highly reliable for the expert to base his decisions upon. To extract the necessary information from cells in the configurations we can either use a point estimate or the interval estimates for the required variables as described in chapter (5) for the 2D case.

Chapter 8

More significant algorithmic changes

Although the methods described in the previous sections to obtain better mixing within a practical time were good, they are still not efficient enough. This is due to the large size of the data set, the overlapping constraint as well as the strong data which favours some configurations above the others. In this chapter we shall introduce some MCMC techniques that are designed to help in exploring the sample space and to achieve a better mixing.

8.1 Simulated tempering

The output samples of the MCMC algorithm after improving it were good, however, the acceptance rates of the moves where less than 15% indicating slow mixing. When the sample space $\mathbb{E}$ has a high dimension such as in our case and the target distribution has unknown isolated modes which are difficult to find and when the chain is slowly mixing, then it may take an astronomically long run of the MCMC algorithm, to get accurate estimates. Our sampler updates approximately one variable at a time so the mixing time increases with the number of variables. To improve the mixing of samplers

of a target model $\pi(x)$ Mariani and Parisi (1992) and Geyer and Thompson (1995) have used simulated tempering. To visit most of the states of the variables they defined a new *temperature* variable T and tried to simulate realisations from a distribution at different temperatures while remaining in equilibrium. This is done by enlarging the configuration state of the system now characterised by two variables: X the original configuration which takes a value on a common state $\mathbb{E}$ and the new discrete variable T which takes values in a specified set. The stationary distribution of the sampler is then

$$\pi(x,t) \propto f(t)\pi_t(x),$$

where $f(t)$ is the prior function of t which is specified in advance and $\pi_t(x)$ is a conditional density of X given the temperature $T = t$. Hence, we have a sequence of densities $\pi_t(x)$ indexed by $T = t$ changing from the coldest temperature, $T = 1$, which is the distribution of interest to the hottest temperature, $T = T^*$, which is hopefully much easier to simulate. For example, $\pi_T(x)$ can be proportional to $\pi(x)^{h(T)}$, where $h(T)$ is an arbitrary function of T; a common function $h(T)$ is $1/T$. The idea is to use these distributions and jump from one state to another such that the algorithm explores different regions of the sample space each time the chain uses the hottest temperature. If the hottest model is chosen such that it is easy to simulate from such as a uniform distribution on $\mathbb{E}$, then the algorithm produces independent draws from that distribution, $\pi(x, T^*)$ and regenerates each time the chain uses this uniform distribution.

The simulating tempering technique treats the index temperature T of the distribution as an additional variable to be updated stochastically. The algorithm updates x using MCMC for $\pi_T(x)$ fixing the temperature at the current $T = t$. It also updates T via a random walk and calculates the Metropolis-Hastings ratio from t to t' for a fixed state x where

$$R = \frac{\pi_{t'}(x)f(t')}{\pi_t(x)f(t)}. \tag{8.1}$$

As equation (8.1) shows, we need to specify the normalising constants for the distri-

butions of the original state variables at the various temperature used, $\int_{x\in\mathbb{E}} \pi_T(x)dx$. If we choose a roughly uniform marginal distribution for T; i.e., the chain spent an equal amount of time at each temperature; then the Pseudo prior $f(t)$ can be used to estimate the normalising constant of $\pi_t(x)$. To avoid inefficiency of a random walk Neal (1996) in his algorithm called *tempered transitions* moves systematically from the desired distribution to the easily sampled one and back to the desired distribution.

To investigate the hottest temperature T^* suitable for our sampler, that is, where we think we get a rapid mixing of the chain, we have simulated samples using $\pi(x, n|y)^{1/t}$ for values of $t = 10, 25, 50, 100, 200$. Our criteria for significant mixing of the sampler is based on the cumulative number of accepted moves during a fixed number of iterations. Table (8.1) lists the total number of accepted moves for 20000 iterations using several values for t.

t	B	Sh	Rs	D	Or	In	M	Sp	Cl
1	33	40	390	4	82	427	0	10	39
10	27	41	346	1	64	424	0	3	29
25	29	42	371	2	67	499	0	4	31
50	30	46	405	3	80	588	0	7	34
100	33	42	404	4	76	626	0	8	37
200	37	47	459	5	88	802	0	9	41

Table 8.1: The total number of accepted birth(B), shift(Sh), resize(Rs), death(D), orientation(Or), updating intensity(In), merge(M), split(Sp) moves and total number of cells(Cl) respectively using $\pi(x, n|y)^{1/t}$.

If we observe the figures in table (8.1) vertically, i.e., within the moves, we find that there is a noticeable difference in the total number of the accepted updating intensity proposals as t increases. For the resize move, although the figures change from one temperature to another, the rate of accepted moves remains approximately the same with regard to the number of the existing cells.

The total number of accepted moves updating the intensity level increases with the values of t. The average accepted move per cell using $t = 1$ is 11% and it increases to 20% using $t = 200$. The acceptance probability for updating the intensity level

move is the only move type which is not affected by the overlapping constraint. This explains the reason beyond the insignificant changes in the total number of accepted moves for the birth, resize, orientation, shift, split and merge moves. The overlapping constraint is strongly influencing the acceptance probability of the moves. It is stronger than the other terms involved in the acceptance probabilities for these moves because it introduce zeros.

Suppose that a higher value of $t > 200$ was found such that the chain mixes rapidly at that temperature, then in order for the simulated tempering algorithm to work efficiently we need to jump between different values of the temperatures ranging from 1 to T^* so often. Thus the values in the sequence have to be close enough that moves between t can be accepted and this would produce far two many steps between $T = 1$ and $T = T^*$.

It looks as if simulated tempering of the type described here will not help much because of the overlapping constraint. What we need instead is a technique to overcome the overlapping constraint and thereby helps us explore different regions of the sample space to achieve the irreducibility of the Markov chain. A relaxed model approach is meant to do that.

8.2 Relaxed model technique

The posterior image distribution given observations $Y = y$ is restricted to scenes in which there is no overlap between cells. The complexity of this non-overlapping constraint along with defining reversible moves, such as merge and split moves, makes it difficult to have efficient move types which maintain the non overlapping constraint and which provide the required mixing of the Markov chain in a reasonable time. For instance, the merge move needs a long time to find two good candidates, say x_A and x_B, to propose to merge them into x_C. The candidates need to be ϵ_d-close and $\epsilon_{\theta_i}-$close for $i = 1, 2, 3$ and $\epsilon_\lambda-$close. With the defined values for the ϵ the probability of finding

such x_A and x_B is very small. If we neglect the overlapping condition and assume that the two cells x_A and x_B are randomly distributed from the prior then the probability of having x_A and x_B with a difference of intensities less than ϵ_λ, that is $pr(|\lambda_A - \lambda_B| \leq \epsilon_\lambda)$ is $(2\epsilon_\lambda)/(\max_\lambda - \min_\lambda)$. This is for the case where each set of properties is an independent realisation from the mark distribution in the reference model (on which the prior is based). Substituting for $\epsilon_\lambda = (\max_\lambda - \min_\lambda)/5$, this probability is 0.4. Using the same assumptions, we compute the probability of having l_A and l_B ϵ_d-close, and it is approximately 3×10^{-5}. It follows that the probability of having two cells satisfying the above five conditions is very small. This probability is even smaller if we consider the non overlapping constraint. The non-overlapping condition not only affects the probability of proposing a viable merge move of two cells, but also the probability of accepting the resulting cell. The chance of proposing, and hence acceptance of, the merge move increases with an increase in the ϵ values, in contrary to the split move where the chance of its acceptance decreases with the increase in the values of the ϵ as we are more likely to propose two overlapping cells. There is a tension between these reverse moves because of the bijection between them. Previous analysis of the sampler results favour the split move over the merge move and, hence, small values of the ϵ are employed. The non overlapping condition is also involved in the acceptance rate of the other MCMC move types. Therefore the resulting Markov chain has difficulty to visit all the region of the sample space $\mathbb{E}$ in practical time.

Geman *et al.* (1990) have dealt with a constraint involved in a model $\pi(x)$ when they aimed to find the value x that maximises $\pi(x)$. The constraint is initially relaxed but they then impose a penalty on the forbidden realisations. This technique could be very fruitful and even essential to achieve irreducibility in situations where the sample space is constrained. An example of reducible Markov Chain is in the genetic pedigree analysis (see Sheehan and Thomas (1993)) when moving from a genotype configuration to another consistent configuration. Suppose we are changing one genotype at a time, then it is not possible to visit some configurations without visiting inconsistent ones.

In these types of problems the need for the relaxed model is crucial to visit some legal samples, and the cost of visiting the illegal cases is high.

The idea of relaxing the constraint is fully explained by Hurn *et al* (1999). They extended the sample space $\mathbb{E}$ to $\mathbb{E}^* \supset \mathbb{E}$ and generate samples from this larger space using $\pi^*(x)$, where $\pi^*(x) > 0, \forall x \in \mathbb{E}^*$ and $\pi(x) = 0, \forall x \in \mathbb{E}^* \backslash \mathbb{E}$. Better mixing is expected for $\mathbb{E}^* \backslash \mathbb{E}$ under $\pi^*(x)$. They described some configurations in $\mathbb{E}^*$ as illegal states in the sense that they do not satisfy the constraint. Using the relaxed model π^* is meant to help in exploring more regions of the sample space by reducing the constraint effects and, hence, achieving the irreducibility of the Markov chain.

In this approach there are two transition kernels for proposed moves, one say $\mathbb{P}$, with stationary distribution the correct distribution $\pi(x)$ and the other $\mathbb{P}^*$ with stationary distribution the relaxed model $\pi^*(x)$. We denote by *legal* chain, the Markov chain which takes values in $\mathbb{E}$ and has limiting distribution π. For one step in this chain, we start with a proposal according to $\mathbb{P}$. This proposal can be an illegal one.

1. If the proposal is legal, the candidate is either accepted or rejected depending on the acceptance probability of the move and that completes the step.

2. If the proposal is illegal, we take steps using kernel $\mathbb{P}^*$ for target distribution π^* on $\mathbb{E}^*$ until an $x' \in \mathbb{E}$ is reached. We then either accept x' and the legal chain will move to x' or reject x' and the legal chain remains where it was before and the step is complete.

Thus the step is complete when either one of the above actions is complete. The steps using $\mathbb{P}^*$ induce acceptance or rejection stage as well.

The acceptance probability of state x' obtained by starting at $x \in \mathbb{E}$ and ending at $x' \in \mathbb{E}$ through some illegal moves in $\mathbb{E}^* \backslash \mathbb{E}$ is

$$\alpha(x, x') = \min\left\{1, \frac{\pi(x')\pi^*(x)}{\pi(x)\pi^*(x')}\right\}. \tag{8.2}$$

To prove the legal chain on $\mathbb{E}$ through some illegal states attains detailed balance

with respect to π, it is sufficient to consider one single path from $x \in \mathbb{E}$ to $x' \in \mathbb{E}$ through some illegal states in $\mathbb{E}^* \backslash \mathbb{E}$ from all available paths. There is a one-to-one correspondence between the set of paths that start in x and end up in x' and the set of paths from x' to x through $\mathbb{E}^* \backslash \mathbb{E}$. Let such a path be $\xi = \{x, x^1, \ldots, x^n, x'\}$. The probability of starting in x and finishing in x' along ξ is

$$P_\xi(x, x') = \mathbb{P}^*(x, x^1)\mathbb{P}^*(x^1, x^2) \ldots \mathbb{P}^*(x^n, x'). \tag{8.3}$$

The secondary MCMC sampler with transition kernel $\mathbb{P}^*$ attains the detailed balance with respect to π^* so

$$\mathbb{P}^*(x^i, x^j) = \mathbb{P}^*(x^j, x^i)\frac{\pi^*(x^j)}{\pi^*(x^i)}, \quad \forall x^i, x^j.$$

Substituting this in equation (8.3), it simplifies to

$$P_\xi(x, x') = \frac{\pi^*(x')}{\pi^*(x)} P_{\xi^-}(x', x), \tag{8.4}$$

where $\xi^- = \{x', x^n, \ldots, x^1, x\}$. Equation (8.4) is true for all paths $\{\xi\}$ so, with $\alpha(x, x')$ as defined in equation (8.2), the detailed balance condition is satisfied for the legal MCMC sampler through some illegal states. As the detailed balance is maintained for the two paths: the one step direct path and via illegal states path, so detailed balance is maintained for the general legal sampler. The probability of moving from one legal state, say x, to another legal state, say x', is the sum of the probability of moving from x to x' via direct step and the probability of moving from x to x' via the second route.

The choice of $\pi^*(x)$ should promote movement between the legal states and give high probability through $\mathbb{E}^* \backslash \mathbb{E}$. This can be achieved by choosing $\pi^*(x)$ proportional to $\pi(x)$ in $\mathbb{E}$ and associated with the weight of illegality of the states in $\mathbb{E}^* \backslash \mathbb{E}$. For our application we start with the reference measure $\Gamma^*(X, N)$ defined on the state space $\mathbb{E} = \cup_n \mathcal{U}^n$ and choose $\mathbb{E}^*$ the same as $\mathbb{E}$. For the prior distribution of (X, N), instead

of using the hard core model with respect to the reference measure, we use another model from the class of pairwise interaction models which will be

$$\pi^*(x,n) = k\beta^{n(x)} \exp(-\gamma \times \text{no. of pixels covered by more than one cell}), \qquad (8.5)$$

where k is a normalising constant computed by integrating the above density over all $(x,n) \in \cup_n \mathcal{U}^n$ with respect to $\Gamma^*(X,N)$, and γ is a penalty rate that would exhibit repulsion or inhibition between the jumps between legal and illegal states. The limiting distribution of π^* as γ tends to ∞ is equivalent to the hard core model. Setting high penalty rates will discourage the chain to be in illegal states whereas using small values for γ increases the chance of moving between the legal and illegal states.

The likelihood function for the records Y will also contribute in defining a relaxed model $\pi^*((X,N)|Y)$ where we shall still use the assumptions stated in section (2.1.2) and modify it to fit the new situation. We assume that the experimental data $Y_j, j = 1,\ldots,S$ are independent given the true scene $X = x$ and normally distributed. If a pixel is covered by more than one cell then we shall set the mean intensity level for that pixel to be the average value of the intensities of the cells covering that pixel, otherwise we set the mean as defined in equation (2.1). The general formula for variance of the records Y remains as defined in (2.1) and corresponds to the changes in the mean values of Y. The relaxed model is then proportional to the product of the prior distribution stated in equation (8.5) and the likelihood.

To reduce the time of moving from one chain to another Hurn *et al.* (1999) used a block technique. Working with a low level (pixel grid) problem they motivated the algorithm by limiting the implementation of this technique to a small block B of component of $\{X\}_B$, where $X = \{X\}_B \bigcup \{X\}_{-B}$ is the output vector. The block B can be chosen randomly or deterministically. They applied the technique using the relaxed distribution π^* only with block B holding the rest of the components $\{X\}_{-B}$ fixed. They used an MCMC sampler with respect to the conditional distribution

$\pi^*(x_B|x_{-B})$ to draw values for the block components. To maintain the irreducibility for the Markov chain the position of the block is randomly selected. Moreover, in order to control the time spent in the illegal states, they have defined a stopping rule to terminate the use of the secondary transition kernel $\mathbb{P}^*$. Detailed balance holds as long as there is one-to-one correspondence in the set of paths from x to x' and from x' to x, and provided detailed balance holds for the chain with respect to $\pi^*(x_B|x_{-B})$. Since $\pi^*(x) = \pi^*(x_B|x_{-B})\pi^*(x_{-B})$, equation (8.2) is seen still to hold.

To investigate the efficiency of this technique for our problem, we apply it using the 2D data. To reduce the size of the illegal state space, we define $\mathbb{E}^*$ to be all the unordered sets of $\{X_1, \ldots, X_N\}$, $X_i \in U$ where at most two cells can cover the same region in L,

$$\pi^*(X, N) \propto \beta^{n(x)} \exp(-\gamma \times \text{no. of pixels covered by two cells}). \tag{8.6}$$

Each 2D cell covers around 900 to 1500 pixels in L and the product of the likelihood terms for these pixels is extremely big. So in order to have an effective relaxed model we ought to have either a high penalty rate γ or instead a higher order function of the overlap terms , i.e., not linear in overlap. As our window L is large we shall use the block-wise approach, where we only allow the illegal states for a small block of the window leaving the other areas of the window unaffected. That is, once the an illegality has been proposed we modify only cells within a small window until resolved. We define the block B to be a sub-window of L with fixed boundary length and completely random location within L. We define X_B to be all the objects with centres lying inside B. Finally to reduce the time spent in illegal states, we shall stop using the secondary (illegal) transition kernel $\mathbb{P}^*$ when the first legal configuration is achieved or when starting from a legal state, the number of iterations spent in the illegal states reaches a fixed number, say m, whichever is smaller. If we spent m steps in the illegal states without hitting a legal configuration then the ending state x^m will be illegal and

γ	LA	IA	IR	SE	AE	UE	CL
0	164	50728	379870	7	160	275	5
50	70	2330	135856	7	154	70	3
150	136	2046	81220	12	247	24	5
250	93	221	4609	10	360	0	6
350	77	320	12139	12	449	3	6
450	113	37	1333	7	26	0	3
550	209	21	1175	7	21	0	5

Table 8.2: The total number of accepted legal moves(LA), accepted illegal moves(IA), rejected illegal moves(IR), number of successful excursions(SE), average length for the successful excursions(AE), number of unsuccessful excursions(UE) and number of existing cells(CL).

$\pi(x') = 0$. That is, the proposal candidate will be rejected and the chain remains in the same legal state it started in.

To look for efficient values of γ we applied the algorithm fixing B to be the sub-window with corner points $(30, 210)$, $(30, 290)$, $(110, 210)$ and $(110, 290)$. We chose the size of B so that it covers more than one cell to have a more effective trial of the relaxed model. The location of B is chosen because of the uncertainty of the objects in that region (see figure $(6 - 1)$). Setting m to 2000 (which was selected based on some pilot tests and considering the overall run-time of the sampler), eight different values of γ were tested. For each value of γ we run the sampler for 2000 legal iterations starting with the empty configuration and allowing up to $m = 2000$ illegal steps between each two legal states. We observed indicators such as the number of legal and illegal moves accepted, the number of rejected illegal moves, number of *excursions* (that is the number of times the chain managed to move through illegal states to legal states in time less than m). We also observed the average length of these successful excursions which took less than m iterations to reach legal state. Finally we observe the number of times the chain reaches the maximum length m without reaching a legal configuration. The results are listed in table (8.2). The total ceiling on the number of total proposals of legal and illegal states is 4000000, however, the actual proposal steps depend on the

illegality states, the excursions length and the initial state. The number of legal states accepted depends on the number of cells in the scene which is stochastic as we started with empty configuration. The figures in table (8.2) indicate that as γ increases, the total number of accepted illegal moves decreases because of the increase in the cost of the illegal states. With a high penalty rate, the expected number of pixels covered by two cells in the accepted illegal configurations is expected to be low in general, therefore the relaxed technique will not be as efficient as required. We also find that the number of unsuccessful excursions is higher using low values of γ. The lengths of the excursions depend on the amount of illegality of the chain, that is they depend on the values of the illegal records and the number of illegal pixels. If the true cells are very close and their intensity level is high then moving to illegal states will be more rapid even for high values of γ (up to a certain limit, e.g. γ greater than 450) and in these situations the successful excursions will be lengthy. The length of the excursions also increases with the number of illegal pixels and as we are proposing small changes for the potential variables involved in the moves we expect that the time needed for the chain to reduce its illegality and move toward the legal space will be long. Therefore the number of unsuccessful excursions increases as γ decreases, which in turn implies an increase in the number of iterations the chain spent in the illegal states and hence, the total number of illegal states rejected. Based on the above criteria we expect that for high values of γ greater than 350, the relaxed model technique will not help significantly. On the other hand, for small values of γ such as γ less than 100, the relaxed model technique will be computationally expensive. However, one might argue that the added computational expenses is acceptable if it helps us to obtain reliable samples. Setting the values of $\gamma = 300$ and 150 we ran the sampler for 30000 legal iterations starting from the empty configuration, using the same fixed block B and limit m, and observed the changes in the output samples. The samples are shown in figures $(8-1)$ and $(8-2)$. Visual inspection of the plots of the consecutive equally spaced MCMC outputs viewed in figure $(8-1)$ when γ is set to 300 suggest no changes

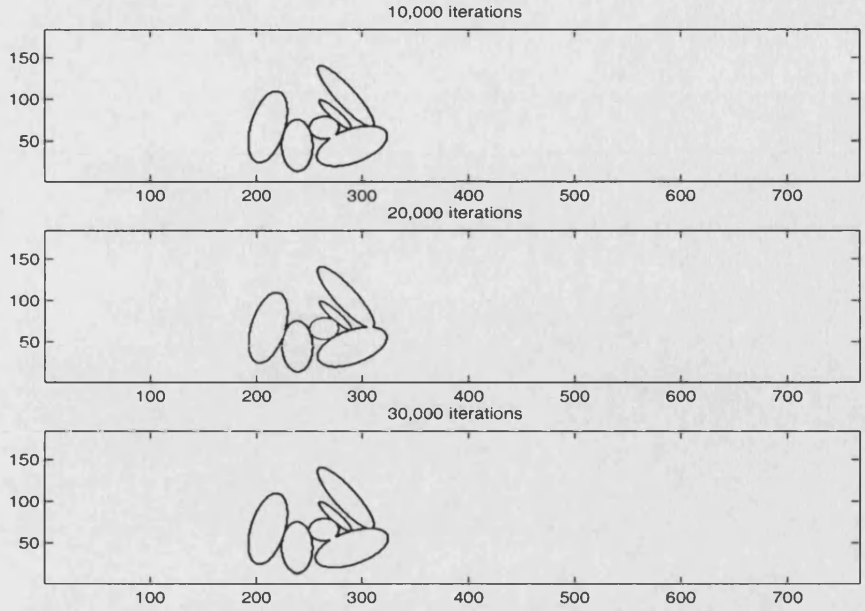

Figure 8-1: The resulting samples after applying the relaxed model technique using $\gamma = 300$ corresponding to iterations number 10000 (top), 20000 (middle) and 30000 iterations (bottom).

in the output configurations, indicating slow mixing of the chain and the inefficient use of the relaxed model technique at that value of γ. On the other hand, when setting $\gamma = 150$ the output samples of the corresponding 10000 and 20000 and 30000 iterations seemed the same and match the data. Therefore we viewed the output sample collected at the 1000 iterations (see figure $(8-2)$) to observe whether any changes had occurred. It is obvious that the sampler managed to adjust the cells into a good fit and overcome the overlapping constraint problem. For the same value of $\gamma = 150$ another run of 20000 iterations using a random block position in L was conducted. The resulting samples corresponding to iterations number $5000, 15000, 30000$ and 45000 are shown in figure $(8-3)$.

Close cells such as those within grids $[30 : 70, 670 : 720]$ and $[50 : 100, 200 : 300]$ managed to change their positions despite the overlapping condition which is a good indication of the benefit of the above technique within again a limited run time.

To apply the relaxed model technique for the 3D problem we start with the efficient

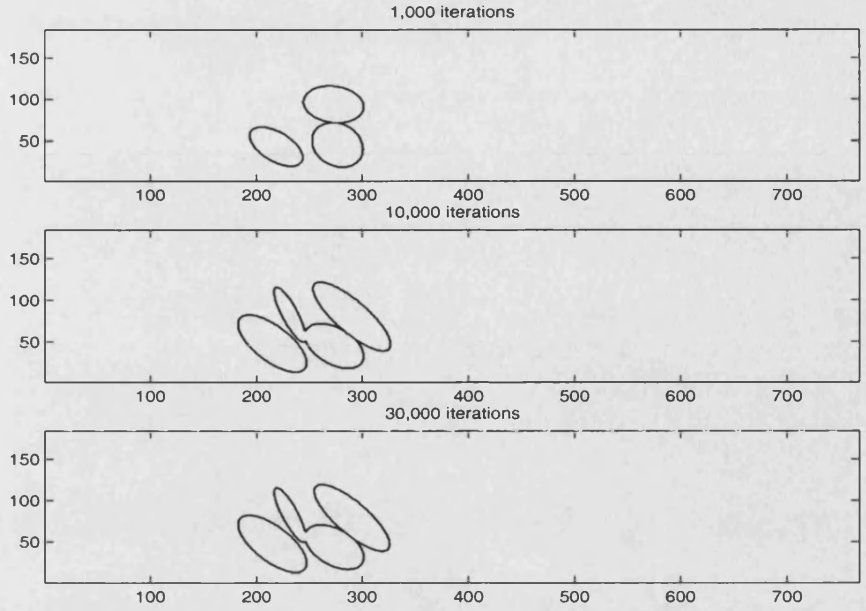

Figure 8-2: The resulting samples after applying the relaxed model technique with $\gamma = 150$ corresponding to iterations number 1000 (top), 10000 (middle) and 30000 iterations (bottom).

values of γ and m obtained in the 2D problem, and applied the technique using a 3D rectangular sub-window of equal length side of 80 voxels, i.e., the block B covers the $z-$axis. Good mixing was achieved for $m = 2000$ and $\gamma = 160$. A sample corresponding to the 5000, 15000 and 25000 iterations starting with configuration with no cells is shown in figures $(8-4)$, $(8-5)$ and $(8-6)$ respectively. The trace of the cells in the three configurations indicates the significant mixing of the sampler and the ability to explore more state space than the previous samplers.

The 2D and the 3D results were significant improvements on previous approaches and the relaxed model technique managed to start from an extreme state and reach the bulk of the posterior image distribution. More significant results would be expected for smaller values of γ and larger values of m. The question which remains is how much sampling effort we are willing to give in order to have a more efficient technique. For the 3D problem the cost is higher because of the dimension of the space and the different type of homogeneity of our window because of the different scales used in the

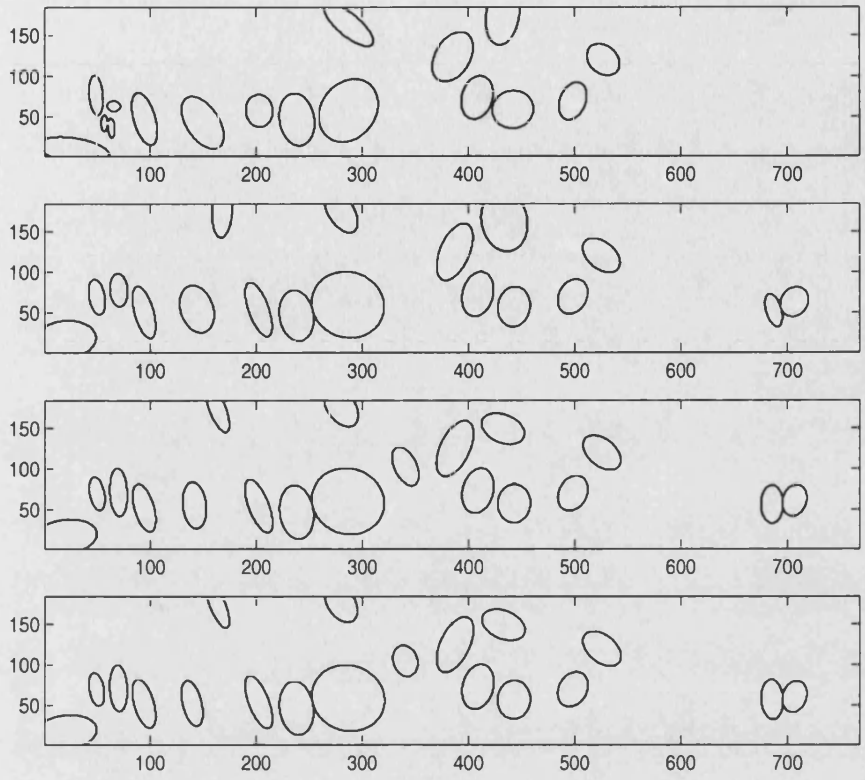

Figure 8-3: The resulting consecutive 2D sections applying the relaxed model technique with $\gamma = 150$ taken at iterations number:(from top to bottom) 5000, 15000, 30000 and 45000.

xy-plane and z-directions. However we think that the relaxed model technique is useful and applicable in this problem.

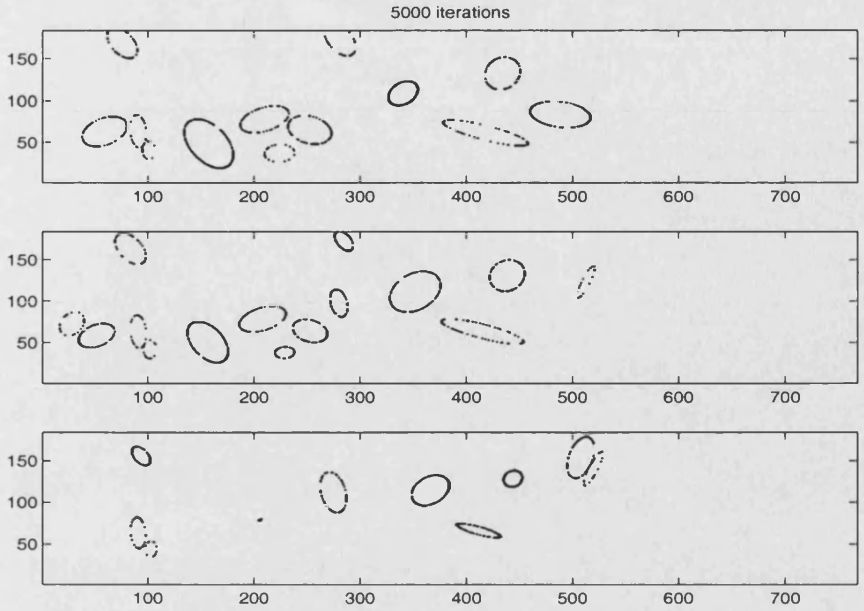

Figure 8-4: The resulting 2D sections at depths z=3, 7 and 12 using the relaxed model technique after running the 3D sampler for 5000 iterations.

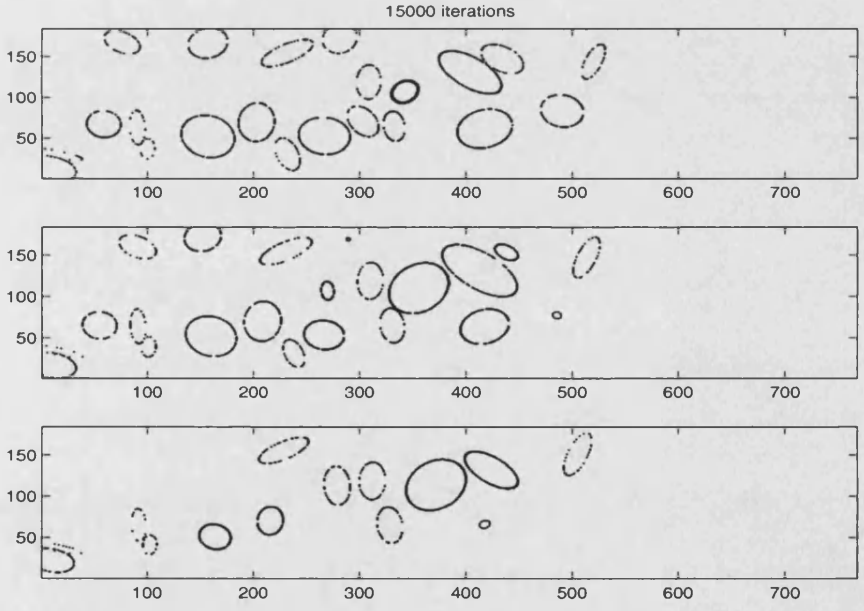

Figure 8-5: The resulting 2D sections at depths z=3, 7 and 12 using the relaxed model technique after running the 3D sampler for 15000 iterations.

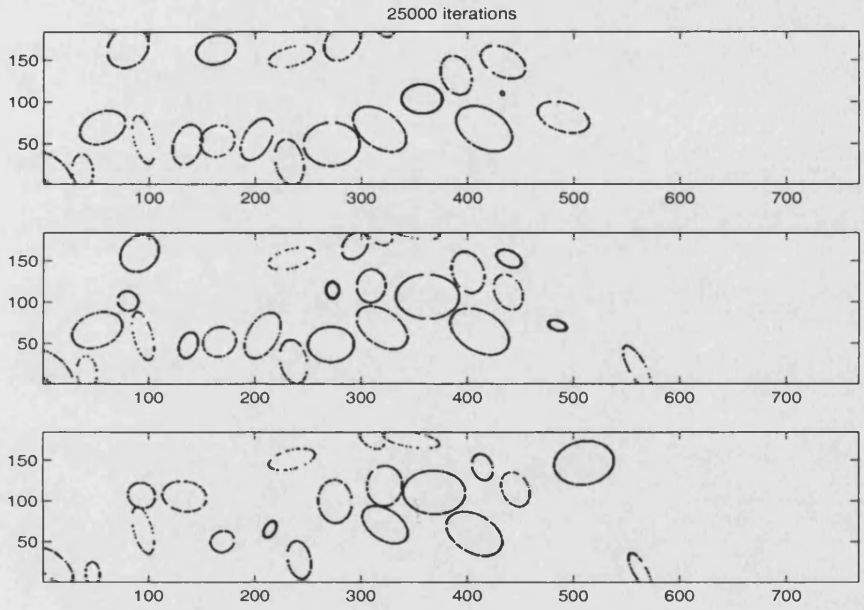

Figure 8-6: The resulting 2D sections at depths z=3, 7 and 12 using the relaxed model technique after running the 3D sampler for 25000 iterations.

Chapter 9

Conclusions and further work

In this thesis we have considered a number of aspects of statistical image recognition. The work has basically fallen into two areas: posterior modelling of the scene and possible improvements to the MCMC reconstruction algorithm.

We worked with data collected using confocal microscopy for an area of cartilage growth. The grey-level image is corrupted by noise, blurring and an attenuation effect. We have provided separate models for two and three dimensional space.

In defining the target models we built the prior models using a marked point process, as was defined by Baddeley and Van Lieshout (1993), to handle the geometric features of the cells. We defined the hard core model as it forbids overlapping among cells in the scene which is a physical structure of a confocal microscope. We have discussed in detail the properties of the data collected from confocal microscopy and expressed these details in an usable likelihood form.

The enormous size of the data structure involved and the complexity of the model made it hard to handle the necessary computations other than by MCMC simulation methods. The space of the configurations is a union of subspaces of different dimensions because the number of cells in the scene was unknown, hence, we used the reversible jump algorithm defined by Green (1995). We used standard and modified Metropolis-Hastings methods to update the configurations of the MCMC chain.

We began the implementation of the chain assuming no information is provided by the user so we started with an empty configuration. We used a visual inspection of the MCMC output and the trace of the total changes in the configurations of the chain as a tool for assessing the convergence and the mixing of our sampler. There was a concern about the convergence of the MCMC algorithms and their ability to explore the sample space, even with substantially more iterations. The very large dimensionality of the parameter space in this problem and the use of single variable updating made monitoring the convergence and the mixing unusually difficult within a reasonable number of iterations. These concerns increased with the existence of the spatial interaction between the cells in the scene which was expressed by the overlapping constraint, a highly influential factor even with lengthy run time.

From a practical perspective, the biologist's interests were on image interpretation and extracting useful information from the data provided based on statistical approaches. Their objective was to estimate and summarise several different features of the cells under study such as their sizes, eccentricity and intensities. Hence, to achieve this objective and avoid prolonged run time we find it is necessary to choose carefully a starting point for the chain near the mode of the posterior image distribution. In the process of constructing initial values we have taken the advantage of operating at different levels of analysis from fine to coarse and vice versa. We applied some morphology operators together with statistical techniques for converting the low level image into a high level valid configuration for our target model. We have also allowed input of expert information to correct obvious errors in the configuration obtained using morphology. These approaches have been applied to both 2D and 3D problems. We spent considerable effort in generalising and applying suitable MCMC techniques for the 3D problem, taking into consideration the finite computational resources.

The issue of a starting point is that the objective of the simulation study is met from the experts point of view and produced a good point estimate for their variables. From our point of view we wanted to find suitable and effective techniques for exploring the

complicated high dimensional uncertainty surfaces such that our output samples will be reliable. We have used independent runs using different seeds for the random number generators used in the MCMC algorithms and collected random samples. Moreover, we ran a single long chain using a relaxed model technique to overcome the difficulties caused by the overlapping constraint. In both cases the key problem was how long the chain should run.

In the 3D problem the drawback of our implementation was first the computational expense of our algorithm, however, this is to be expected because of the large sample size and the huge data sets of the 3D case. The work of improving the computational implementation of the MCMC algorithm remains an area for continued work. The second drawback concerns the choice of the parameters involved in the likelihood form. The model was highly sensitive to some of these parameters such as those involved in defining the relation between the mean and the variance of the cells. Our default settings for these parameters were mainly extracted from the visual analysis of the data provided. However, there is always the fully Bayesian approach to handle these parameters treating them as hyper-parameters in a hierarchical model and assigning them priors. This approach increases the computational time especially with the need for the normalising constants of the model conditional on values of these parameters. Though, one might argue that the added computation expenses are acceptable. Also different relations between the mean and the variance of the cells rather than the linear one we used may be tested, or even a naive model assuming a common intensity level may be applied.

The main properties of the 3D records obtained by confocal microscopes are attenuation, blurring and noise. These features interact together to produce the final data. We had deal with each factor separately. For example, the blurring as illustrated by White *et al.* (1995) is affected by the abberations of the light that hits the focal point and is then transmitted to the detector. That is the blurring is influenced by the depth and the regions that the light travels through. This affects the blurring in the 3D data

and causes changes in the point spread functions through vertical sections. We have assumed no changes in the blurring effect along the consecutive sections. In future work we may be more precise in assessing the blurring effects using deconvolution forms in addition to some other technical details of the changes in the angles and frequencies of the incoming light and outgoing fluorescent beams in the confocal microscopes. This is such a rich area to investigate.

Further investigations to assess the amount of effort spent to get reliable estimates can be carried out in future using a higher level of interaction with the expert. Using expert knowledge and his/her experience in providing reliable information regarding the variables of interest will certainly allow us to assess the simulated estimators and the accuracy of these estimators under each model. This helps to compare the amount of effort spent and the reliability of the results. It also gives an indication about the length of the Markov chain required to reach convergence. To assess the difference between two output samples instead of using visual inspection we can use Baddeley's *delta metric* for binary images (Baddeley(1992)). To use this approach we have to work with pixel level and obtain MAP estimate (Rue and Hurn (1997)) classifying each pixel in the grid as a background or foreground (cells) media. This approach can be used to produce better estimates of the true outline of each object conditioned on the number of objects.

Another area of work which can be achieved in the future is the existence of spatial variation in the observation window. For example, we know that because of the structuring of the cartilage, cells that are close tend to be of same type and, hence, have mainly the same distribution of properties. This spatial pattern of the regions of the specimen may be handled by more sophisticated spatial modelling techniques.

The demand for statistical analysis tools and techniques in imaging has increased substantially in recent years. Although we had built our models and developed the algorithms specifically for this set of data, the implications of our study go beyond that. The methods we have developed are applicable for any set of data collected by

confocal microscopy, which is now a very commonly used imaging device as it images the 3D specimen without the need for physical sectioning. We may need to apply some slight modifications of the model and the MCMC algorithm such as defining different object identification properties and different object spaces. Or we may need to use alternative characteristics such as colour or even surface texture of the objects. A more general application may cover more than one type of standard structure of the objects' space or even deformable structure shape objects. Fourier transformation concepts can be used to define the reference points on the surface of the objects as identifying points (see Hurn (1998)). The task of object recognition is a very wide and demanding area. It is applicable in different fields such as satellite images, geological scans and medical fields. We think that this study provides the user with the basic concepts and skills needed to handle different applications in image analysis or even with other fields using Bayesian inference and MCMC simulations.

Bibliography

[1] Adam, R. and Bischot, L.(1994), Seeded Region Growing, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16(6), 640-647.

[2] Baddeley, A.(1992), An error metric for binary images, In: *Robust Computer Vision: Quality of Vision Algorithms*, 59-78, (editors W. Forstner and S. Ruwiedel). Wichmann Verlag.

[3] Baddeley, A. and Van Lieshout, M.(1993), Stochastic Geometry Models in High-level Vision , *Statistics and Images*, Volume 20, Chapter 11, 235-256, (editors K.V. Mardia and G.K. Kanji). Abingdon: Carfax Publishing.

[4] Besag, J.(1974), Spatial Interaction and the Statistical Analysis of Lattice Systems, *Journal of the Royal Statistical Society*, Series B, 36(2), 192-236.

[5] Besag, J.(1986), On the Statistical Analysis of Dirty Pictures, *Journal of the Royal Statistical Society*, Series B, 48(3), 259-302.

[6] Besag, J.(1989), Towards Bayesian Image Analysis, *Journal of applied statistics*, 16(3), 395-407.

[7] Besag, J., York, J. and Mollie, A.(1991), Bayesian Image Restoration with Two Application in Spatial Statistics, *Annals of the Institute of Statistical Mathematics*, 43, 1-59.

[8] Besag, J. and Green P.J.(1993), Spatial Statistics and Bayesian computation, *Journal of the Royal Statistical Society*, Series B, 55, 25-37.

[9] Besag, J., Green, P., Higdon, D. and Mengersen, K.(1995), Bayesian Computation and Stochastic Systems, *Statistical Science*, 10, 3-66.

[10] Brooks, S.P. and Giudici, P.(1998), Diagnosing Convergence of Reversible Jump MCMC Algorithm. *Bayesian Statistics*, 6, 733-742.

[11] Burkitt, H.G., Young, B. and Health, J.W.(1993), Wheaters Functional Histology. London: Churchill Livingston.

[12] Cappe, O. Robert, C.P. and Ryden, T. (2001), Birth and Death Jump Processes for Inference in Hidden Markov Models, *Technical Report*, University Paris IX.

[13] Canny, J.(1986), A Computational Approach to Edge Detection, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-8(6), 679-698.

[14] Diggle, P.J.(1983), Statistical Analysis of Spatial Point patterns. London: Academic press.

[15] Frigessi, A. and Stander, J.(1994), Informative Priors for the Bayesian Classification of Satellite Images, *Journal of the American Statistical Association*, 89, 426, 703-709.

[16] Geman, S. and Geman, D.(1984), Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-6, 721-741.

[17] Geman, S., Geman, D., Graffigne, C. and Dong, P.(1990), Boundary Detection by Constrained Optimisation, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12, 609-628.

[18] Geyer, C.J.(1991), Markov Chain Monte Carlo Maximum Likelihood, *Computer Science and Statistics*, 23, 156-163. Markov chain maximum likelihood for dependent data, *Journal of the Royal Statistical Society*,B, 54, 657-699.

[19] Geyer, C.J. and Moller, J.(1994), Simulation Procedures and Likelihood Inference for Spatial Point process, *Scandanavian Journal of Statistics*,21, 359-374.

[20] Geyer, C.J. and Thompson, E.A.(1995), Annealing Markov Chain Monte Carlo with Applications to Ancestral Inference, *Journal of the American Statistical Association*, 90(431), 909-920.

[21] Geyer, C.J.(1996), Likelihood Inference for Spatial Point Processes, *in* W.S.Kendall (ed.), *Proc. Seminaire Europeen de Statistique Tolouse 1996, Stochastic Geometry: Theory and Applications*, Springer Lecture Notes.

[22] Gilks, W., Richardson, S. and Speigelhalter D.(1996), Markov Chain Monte Carlo in Practice. London: Chapman and Hall.

[23] Glasbey, C. and Horgan, G.(1995), Image Analysis for the Biological Sciences. Chichester: John Wiley.

[24] Green, P.J. and Han, X.(1992), Metropolis Methods, Gaussian Proposals and Antithetic Variables. In: Stochastic Models, Statistical Methods and Algorithms in Image Analysis *Lecture notes in statistics*, 74, 142-164. Berlin: springer-Verlag.

[25] Green, P.J.(1994), Discussion of Paper by U. Grenander and M. Miller, *Journal of the Royal Statistical Society*, Series B, 56, 589-590.

[26] Green, P.J.(1995), Reversible jump Markov Chain Monte Carlo Computation and Bayesian Model Determination, *Biometrika*, 82(4), 711-732.

[27] Grenander, U. and Keenan, D.M.(1993), Towards Automated image Understanding, *Advances in Applied Statistics*, 16, 207-221.

[28] Grenander, U. and Miller, M.I.(1994), Representations of Knowledge in Complex Systems, *Journal of the Royal Statistical Society*, Series B, 56, 549-603.

[29] Hastings, W.(1970), Monte Carlo Sampling Methods using Markov Chains and their Applications, *Biometrika*, 57, 97-109.

[30] Hurn, M. and Jennison, C.(1993), Multiple Site Updates in Maximum A Posteriori and Marginal Posterior Modes Image Estimation, In: *Statistics and Images*, Volume 1, 155-186, (editors K.V. Mardia and G.K. Kanji). Oxford: Carfax Publishing Company.

[31] Hurn, M. and Rue, H.(1997), High level Image Priors in Confocal Microscopy Applications,In: *The Art and Science of Bayesian Image Analysis* , 36-43, (editors K.V. Mardia, C.A. Gill and R.G. Aykroyd). Leeds: Leeds University Press.

[32] Hurn, M.(1998), Confocal Fluorescence Microscopy of Leaf Cells : An Applications of Bayesian Image Analysis, *Applied Statistics, Journal of the Royal Statistical Society*, Series C, 47, 361-377.

[33] Hurn, M., Rue, H. and Sheehan., N.A.(1999), Block Updating in Constrained Markov Chain Monte Carlo Sampling, *Statistics and Probability Letters*, 41, 353-361.

[34] Kingman, J.F.C. and Taylor, S.J.(1966), Introduction to Measure and Probability. Cambridge: Cambridge University Press.

[35] Marinari, E. and Parisi, G.(1992), Simulated tempering: A New Monte Carlo Scheme. *Europhysics letters*, 19(6), 451-458.

[36] Metropolis, N., Rosenbluth, A., Rosenbluth, M., Teller, A. and Teller, E.(1953), Equations of State Calculations by Fast Computing Machines, *Journal of chemical physics*, 21, 1087-1092.

[37] Moller, J.(1989), On the Rate of Convergence of Spatial Birth-and-Death Processes, *Annals of the Institute of Statistical Mathematics*, 41, 565-581.

[38] Neal, R.(1996), Sampling from Multimodal Distributions using Tempered Transitions, *Statistics and Computing*, 6, 353-366.

[39] Noda, M.(1993), Cellular and Molecular Biology of Bone. USA: Academic Press.

[40] Nummelin, E.(1984), General Irreducible Markov Chains and Non-Negative Operators. Cambridge: Cambridge University Press.

[41] Pawley, J.B.(1996), Handbook of Biological Confocal Microscopy. London: Plenum Press.

[42] Preston, C.(1977), Spatial Birth-and-Death Processes, *Bulletin of the International Statistical Institute*,Book2, 46, 371-391.

[43] Richardson, S. and Green, P.J.(1997), Bayesian Analysis with Mixtures with an Unknown Number of Components (with discussion), *Journal of the Royal Statistical Society*, Series B, 59, 71-80.

[44] Ripley, B.(1981), Spatial Statistics. New York: John Wiley.

[45] Ripley, B.(1987), Stochastic Simulation. New York: John Wiley.

[46] Ripley, B.(1988), Statistical Inference for Spatial Processes. Cambridge: Cambridge University Press.

[47] Rue, H. and Hurn, M.(1997), Loss Functions for Bayesian Image Analysis, In: *The Art and Science of Bayesian Image Analysis*, 72-79, (editors K.V. Mardia, C.A. Gill and R.G. Aykroyd). Leeds: Leeds University Press.

[48] Rue, H. and Hurn, M.(1999), Bayesian Object Identification, *Biometrika*, 86, 649-660.

[49] Ruskov, D.A.(1993), Estimation of Size Distribution of Closed Cell Elements from Analysis of their Random Plane Section, *Biometrics*, 49, 141-149.

[50] Serra, J.(1982), Image Analysis and Mathematical Morphology. London: Academic Press.

[51] Serra, J.(1986), Introduction to Mathematical Morphology, *Computer Vision, Graphics and Image Processing*, 35, 283-305.

[52] Serra, J.(1988), Image Analysis and Mathematical Morphology. Volume 2. London: Academic Press.

[53] Shaw, P.J. and Rawlins, D.J.(1991), Three Dimensional Fluorescent Microscopy, *Progress in Biophysics and Molecular Biology*, 56, 187-213.

[54] Shaw, P.J. and Rawlins, D.J.(1991), The Point Spread Function of a Confocal Microscope: its Measurement and Use in Deconvolution of the 3-D Data, *Journal of Microscopy*, 163, 151-165.

[55] Sheehan, N. and Thomas, A.(1993), On the Irreducibility of Markov Chain Defined on a Space of a Genotype Configurations by a Sampling Technique, *Biometrics*, 49, 163-175.

[56] Smith, A.F.M. and Roberts, G.O.(1993), Bayesian Computation via the Gibbs Sampler and Related MCMC Methods, *Journal of the Royal Statistical Society*, Series B, 55, 3-23.

[57] Speigel, M.R.(1967), Schaum's Outline of Theory and Theoretical Mechanics with an Introduction to Lagrange's Equations and Hamiltonian Theory. New York: MacGraw-Hill.

[58] Tatsuoka, M.M.(1971), Multivariate Analysis: Techniques for Educational and Psychological Research. New York: John Wiley.

[59] Tierney, L.(1994), Markov Chain for Exploring Posterior Distributions, *Annals of Statistics*, 22, 1701-1728.

[60] Tierney, L.(1998), A Notes on Metropolis-Hastings Kernels for General State Spaces, *Annals of Applied Probability*, 8, 1-19.

[61] Van Lieshout, M.N.(1995), Markov Point Processes and their Applications in High-level Imaging, *Bulletin of the International Statistical Institute*,Book2, 559-576.

[62] Wilson, T.(1990), Confocal Microscopy. London: Academic press.

[63] White, N.S., Errington, R.J., Fricker M.D. and Wood, J.L.(1996), Aberration Control in Quantitative Imaging of Botanical Specimens by Multidimensional Fluorescence Microscopy,*Journal of Microscopy*, 181(2), 99-116.

[64] Woessner, F., David, Jr. and Howell, S.(1993), Joint Cartilage Degradation, Basic and Clinical Aspects. USA: Marcell Dekker Inc.